Maximum likelihood estimation and confidence bands for a discrete log-concave distribution

Fadoua Balabdaoui[1]*, Hanna Jankowski[2], and Kaspar Rufibach[3]

July 21, 2011

[1]CEREMADE, Université Paris-Dauphine, Paris, France
[2]Department of Mathematics and Statistics, York University, Toronto, Canada
[3]Division of Biostatistics, Institute for Social and Preventive Medicine, University of Zurich, Switzerland

Abstract

The assumption of log-concavity is an attractive and flexible nonparametric shape constraint in distribution modelling. In this work, we study the maximum likelihood estimator (MLE) of a log-concave probability mass function. We show that the MLE is strongly consistent and derive pointwise asymptotic theory, which is used to calculate confidence bands for the true probability mass function. The proposed estimator and associated confidence bands may be easily computed using the R package logcondiscr. We illustrate the flexibility of the estimator on recent data from the H1N1 pandemic in Ontario, Canada.

Keywords: nonparametric estimation, shape-constraints, confidence interval, H1N1, discrete distribution, Gaussian process, log-concave distribution, envelope

1 Introduction

Nonparametric maximum likelihood estimation of a log–concave probability density in the continuous setting has attracted considerable attention over the last few years. The list of references is extensive, and we refer the reader to Walther (2009); Cule et al. (2010); Seregin and Wellner (2010) and the references therein for an overview of recent theoretical and computational developments. The merits of using log–concavity as a shape constraint have been discussed in detail in Balabdaoui et al. (2009), Cule et al. (2010), Walther (2009), and Dümbgen and Rufibach (2011). Not only do many parametric models admit a log–concave density, but log–concavity is also a valuable surrogate for unimodality. Indeed, log–concavity implies unimodality, and whereas the nonparametric MLE of a unimodal density does not exist (see e.g. Birgé, 1997), it does exist under the assumption of log-concavity. Furthermore, a log-concave density is the natural generalization of the normal density, where $\log f(x) = -x^2/2$,

*Corresponding author. Email address: fadoua@ceremade.dauphine.fr

the "canonical" concave function. However, log–concavity gives more flexibility in modelling since it allows for asymmetry. It also allows for a wide-range of tail-behaviour, with the slowest being that of the double-exponential where $\log f(x) = -|x|$.

Given the large corpus of work on estimation of a log–concave density in the continuous setting, it comes as a surprise that little attention has been given to estimation of a log–concave probability mass function (pmf). The unpublished Master's thesis of Weyermann (2008) is the only work of which we are aware. As in the continuous setting, the assumption of log-concavity is highly appealing. Many discrete parametric models admit a log-concave pmf. Binomial, negative binomial, geometric, hypergeometric, uniform, Poisson, hyper-Poisson (Bardwell and Crow, 1964; Crow and Bardwell, 1965), the Pólya-Eggenberger, and the Skellam distribution (Karlis and Ntzoufras, 2006; Alzaid and Omair, 2010) are some examples; see Johnson and Kotz (1969) and Devroye (1987) for further details. Therefore, the log-concave assumption provides a broad and flexible, yet natural, non-parametric class of distributions on $\mathbb{Z}$. In Section 2.1, we give a formal definition of log–concavity of a pmf. The motivation behind this formalism is to make the link between the different definitions given in the literature, and also to avoid some pitfalls as illustrated in a counterexample given in (2.1). In addition, as we show in Section 2.2, if data admitting a continuous log-concave density has been grouped, then the resulting pmf is also log-concave. Therefore, the log-concave pmf provides an appropriate surrogate for many data problems where the log-concave density assumption would have been used, but where the data has been grouped or discretised.

In Figure 1, we show an example of the nonparametric MLE under the log-concave assumption. The true underlying distribution is the negative binomial with parameters $r = 6$ and $p = 0.3$. The behaviour of the MLE is considerably better than that of the empirical mass function; the MLE is able to smooth out the rather erratic empirical pmf and recover the general shape of the true pmf quite well, even for these small sample sizes. Further simulations are shown in Section 5, where we study the finite sample performance of the proposed estimator via simulations. Here, we find that the MLE performs well, even when compared to the parametric estimator, but more so when the logarithm of the true pmf does not have steep slopes. This behaviour can be further explained by the theoretical results of Section 4.

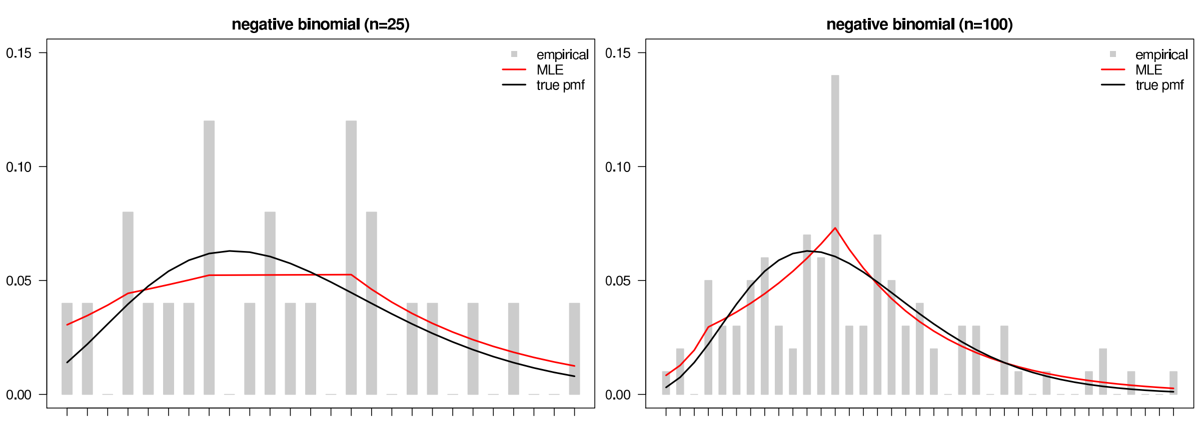

Figure 1: Nonparametric MLE of the negative binomial (6, 0.3) distribution for $n = 25$ (left) and $n = 100$ (right).

In Weyermann (2008), it was shown that the MLE of a log-concave pmf on $\mathbb{Z}$ exists and is

unique. Furthermore, the logarithm of the MLE is piecewise linear and we use the term knot point to describe any point where a piecewise linear function has a bend. As in the continuous case, the MLE admits knots at some subset of the set of observations. Weyermann (2008) also provides an active set algorithm to compute the MLE, much in the spirit of Dümbgen et al. (2010), which was implemented in Matlab. We have adapted this code to R (R Development Core Team, 2011) in the new package **logcondiscr** (Rufibach et al., 2011), available from CRAN. In Section 3, we recall some of the main results of Weyermann (2008), and provide additional characterisations of the estimator. These characterisations are important in the study of the estimator and provide crucial insight into its behaviour. In this section, we also give some details on the computation of the MLE.

In Section 4, we study the asymptotic behaviour of the MLE. Our main theoretical results establish strong consistency and robustness of the MLE, as well as its pointwise weak convergence on any segment between two successive knot points of the logarithm of the true pmf which are a finite distance apart. For our theory of weak convergence to hold, we require that the true log-concave pmf admits a one-sided support. Note that this assumption is not too restrictive as it is satisfied by most of the known parametric log-concave discrete distributions. Unlike the continuous setting in $\mathbb{R}$, which admits rates of $n^{2/5}$ for a twice differentiable density, here convergence occurs at the parametric rate $\sqrt{n}$. The limit distributions of the MLE, its logarithm, and the corresponding cumulative distribution function (cdf) are explicitly given. These limits are characterised in terms of an envelope-type process, $\mathbb{H}$, which can be viewed as a discrete analogue of the random process of Balabdaoui et al. (2009), also appearing in the pointwise convergence of the MLE of a convex decreasing density; see Groeneboom et al. (2001b). We also show that the limiting process can be described as the solution of an appropriate least squares concave regression problem. The advantage of this approach is that the solution of the least squares problem can be found explicitly using, for example, the R package **cobs** (Ng and Maechler, 2011). Therefore, we are able to sample directly from the limiting distribution, which allows us to compute pointwise confidence bands for the true pmf. The details of our approach are given in Section 4.4.

To our best knowledge, this is for the first time that confidence bands have been computed in the log-concave setting. In the continuous case, Balabdaoui et al. (2009) derive pointwise asymptotics for the MLE. However, these depend on the value of $\psi''(x_0)$, where $\psi = \log f$, which is difficult to estimate. Similar problems arise for the monotone shape-restriction, and the work of Banerjee and Wellner (2001) was designed to overcome this issue. Currently, no such results exist for the log-concave MLE on $\mathbb{R}$.

As an illustration of our methods, we apply the proposed estimator to a real data set of H1N1 influenza pandemic data from Ontario, Canada. This data comes from an early study of the pandemic, Tuite et al. (2010), when it was crucial to provide a quick analysis of the behaviour of the virus. The flexibility of the log–concave assumption makes it suitable to describe important aspects of the incubation period of the swine flu, as well as the duration of symptoms. To handle errors in the data collection, we use a simple mixture model, which gives even more flexibility to our approach.

Conclusions and a discussion can be found in Section 7. Appendix A provides some notation which is used throughout this paper, and Appendix B discusses some of the properties of log-concavity. There, we also give some classical examples of pmfs for which we verify log-

concavity and give the explicit location of the corresponding mode when possible. All proofs and technical details have been collected in Appendix C.

2 Log–concavity of discrete distributions

Our first goal is to give a precise definition of log–concavity in the discrete setting. Although such a property might seem to be obvious to describe mathematically, we find that it is necessary to give a formal definition to avoid certain pitfalls. Some further details of log-concavity are provided in Appendix B. In particular, we give a definition of log–concavity based on that of concavity. To our surprise, no reference of concavity in the discrete setting is available, and hence our effort to fill this gap in Section 7.1. Beyond what is provided below, we refer the reader to Keilson and Gerber (1971) and Dharmadhikari and Joag-Dev (1988) for further details on log-concavity in the discrete setting.

2.1 Definition of log–concavity

Recall that the definition of log-concavity is closely tied with that of unimodality. This is true for densities as well as probability mass functions. Consider then a pmf $p = \{p_k, k \in \mathbb{Z}\}$.

Definition 2.1. *(Unimodality) There exists at least one $k_m \in \mathbb{Z}$ such that*

$$\begin{aligned} p_k &\geq p_{k-1} \quad \text{for all } k \leq k_m \\ p_{k+1} &\leq p_k \quad \text{for all } k \geq k_m. \end{aligned}$$

The integer k_m in the definition above does not have to be unique. In case there exist several integers satisfying the same property, the smallest one will be defined as the mode of p.

Definition 2.2. *(Strong unimodality) p is strongly unimodal if for any unimodal pmf q, the convolution $p \star q$ is again a unimodal pmf.*

As indicated by Keilson and Gerber (1971) strong unimodality implies unimodality since the Dirac distribution at zero is unimodal. In their Theorem 3, they prove equivalence between strong unimodality and log–concavity.

Definition 2.3. *(Log–concavity) The pmf p is log-concave if both of the following conditions hold:*

A. *If $i < j < k$ are such that $p_i \, p_k > 0$, then $p_j > 0$.*

B. $p_k^2 \geq p_{k-1} p_{k+1}$ *for all $k \in \mathbb{Z}$.*

The definitions above extend of course to nonnegative functions which are not necessarily probability mass functions. In Definition 2.3, Property A says that the support $\mathcal{S}$ of p is of the form $\{k_1, k_1 + 1, ..., k_2 - 1, k_2\}$ where $k_1 \leq k_2$, $k_1 \in \mathbb{Z} \cup \{-\infty\}$ and $k_2 \in \mathbb{Z} \cup \{+\infty\}$. Other equivalent descriptions were proposed in the literature. For instance, Keilson and Gerber (1971) express the same property in terms of connectivity (equivalent here to convexity) of the support of p. The second description is due to Stanley (1989) which requires that the

sequence $\{..., p_{-1}, p_0, p_1, ...\}$ does not have an *internal zero.* Property A is far from being superfluous, as illustrated in the following example, given to us by Marios Pavlides in a private communication. Consider the pmf p such that

$$p_0 = 1/2, p_1 = p_2 = 0, p_3 = 1/2, \text{ and } p_k = 0 \text{ for } k \in \mathbb{Z} \setminus \{0, 1, 2, 3\}. \tag{2.1}$$

Although p does satisfy Property B, it is easy to see that it is *not* unimodal, and hence it is *not* strongly unimodal or equivalently *not* log-concave. In Devroye (1987), a definition of log-concavity based on Property B only is given. Unless it is explicitly said that the support of the log-concave sequence is equal to $\mathbb{Z}$, the counterexample in (2.1) indicates clearly that Property B alone is not enough.

The next proposition gives an equivalent condition for log–concavity of a pmf p.

Proposition 2.1. *A pmf $p = \{p_k, k \in \mathbb{Z}\}$, with support $\mathcal{S}$, is log-concave if and only if $\mathcal{S}$ is a subset in $\mathbb{Z}$ of consecutive integers and the sequence $\{p_k/p_{k-1}, k \in \mathcal{S}\}$ is nonincreasing. Here, the first term of the ratios is allowed to take the value ∞.*

Note that the proposition above gives the possibility of constructing an estimator based on "monotonising" the empirical probability ratios. This alternative approach will be pursued elsewhere.

One may also discuss log-concavity in terms of the "slopes" of the function $\psi(k) = \log p_k$. For any $k \in \mathcal{S}$, $(\Delta\psi)_k \leq 0$ for a log-concave pmf, where $(\Delta\psi)_k = \psi_{k+1} - 2\psi_k + \psi_{k-1}$ denotes the discrete Laplacian. However, the condition $(\Delta\psi)_k \leq 0$ for $k \in \mathbb{Z}$ is not sufficient to check for log-concavity of a pmf, as $(\Delta\psi)_k$ may not be well-defined. For this reason, we provide more exact conditions on concavity in Appendix B.

Definition 2.4. *For any concave function φ with $\mathcal{S} = \{k \in \mathbb{Z} : \varphi(k) > -\infty\}$, we call any point $k \in \mathcal{S}$ such that $(\Delta\varphi)_k < 0$ a knot point of the function φ. Here, we use the convention $a - \infty < 0$ for any real number a. Furthermore, if both k and $k+1$ are knots, then we say that k is a double knot, and if $k-1, k$ and $k+1$ are knots, then we say that k is a triple knot.*

2.2 Grouped versus continuous data

In this section, we establish the link between continuous and discrete log–concave distributions. Let $\{A_i, i \in \mathbb{Z}\}$ denote a partition of of the positive real line such that $|A_i|$ is constant. We assume that each interval A_i is either of the form $(\alpha_i, \beta_i]$ or $[\alpha_i, \beta_i)$. That is, we assume that each interval is either right–open and left–closed, or vice versa. For a continuous random variable X admitting a density we then define the probability mass function $p_i = P(X \in A_i)$. Such a scenario arises either under grouping (as in the H1N1 example considered in Section 6) or discretisation, where one observes $Y = \delta\lfloor X/\delta \rfloor$, for example, instead of the continuous random variable X. We have the following result.

Proposition 2.2. *Suppose that the continuous random variable X has a log-concave density with respect to Lebesgue measure on $\mathbb{R}$. Then the probability mass function $p = \{p_i\}$ is also log-concave.*

Throughout this paper, we focus on the probability mass function defined on $\mathbb{Z}$. However, our results are applicable to a pmf defined on any regular grid, as long as that grid does not depend on sample size, as was considered in Tang et al. (2011).

3 Properties of the maximum likelihood estimator

3.1 Existence

Let $(X_1, \ldots, X_n)$ denote a random sample from the pmf p. Then, the MLE of a log-concave pmf is found by maximising the log-likelihood $\sum_{i=1}^n \log p_{X_i}/n$, over the class of log-concave pmfs, $\mathcal{LC}_1$. Let m be the total number of the ordered distinct values $k_1 < \cdots < k_m$ occurring in the sample $(X_1, ..., X_n)$, $\mathcal{I} = \{k_1, \ldots, k_m\}$ the set of these observations, and $w_j = \bar{p}_{n,k_j} = n^{-1}\sum_{i=1}^n 1_{\{X_i = k_j\}}$, $j = 1, ..., m$, the corresponding empirical frequencies. By Theorem 3.1 of Silverman (1982), the MLE can also be found as the maximiser of $\sum_{j=1}^m w_j \log(p_{k_j}) - \sum_{k\in\mathbb{Z}} p_k$, over the class $\mathcal{LC}$ (the class of log-concave nonnegative sequences), where the term $\sum_{k\in\mathbb{Z}} p_k$ is the Lagrange multiplier. Using the one-to-one correspondence between the classes $\mathcal{LC}$ and the class of all concave functions $\mathcal{C}$, it follows that the MLE exists if and only if the criterion function

$$\Phi_n(\psi) = \sum_{j=1}^m w_j \psi(k_j) - \sum_{k\in\mathbb{Z}} \exp \psi(k) \tag{3.1}$$

admits a maximiser $\widehat{\psi}_n$ over $\mathcal{C}$. Then, the maximum likelihood estimator is given by $\widehat{p}_{n,k} = \exp \widehat{\psi}_n(k)$ for $k \in \mathbb{Z}$.

Reducing the set of functions over which Φ_n is maximised is one of the key steps in proving existence of $\widehat{\psi}_n$. It also sheds more light on the shape of the estimator, and is of crucial importance when setting up an algorithm to compute $\widehat{p}_n$. We first introduce the family $\mathcal{F}_m$ of functions

$$\mathcal{F}_m \;:=\; \{\varphi \;:\; \mathbb{Z}\cap[k_1, k_m] \to \mathbb{R},\ \varphi = -\infty \text{ on } \mathbb{Z}\cap\{\mathbb{R}\setminus[k_1, k_m]\}\}.$$

For any $\varphi \in \mathcal{F}_m$, we consider the set of knot points $\mathcal{K}(\varphi) = \{k \in \mathbb{Z}\cap[k_1, k_m] \;:\; (\nabla\varphi)_k < 0\}$. Note that k_1 and k_m are always in $\mathcal{K}(\varphi)$. If we consider the piecewise function defined on $[k_1, k_m]$ interpolating φ between the points $k_j, j = 1, ..., m$, then the knot points of φ are exactly the bending points of its interpolating function; i.e., the points at which this function changes its slope. Finally, we consider the sub-family $\mathcal{F}_m(\mathcal{I}) = \{\varphi \in \mathcal{F}_m \;:\; \mathcal{K}(\varphi) \subseteq \mathcal{I}\}$ of functions in $\mathcal{F}_m$ which only admit knots in the set of observations $\mathcal{I} = \{k_1, ..., k_m\}$, and we let $\mathcal{C}_m(\mathcal{I})$ be the subset of concave functions φ in $\mathcal{F}_m(\mathcal{I})$.

The following results, which we recall here for completeness of exposition, have been shown in the unpublished Master's thesis Weyermann (2008). A shortened proof is given in Appendix C.

Theorem 3.1 (Weyermann, 2008)**.** *We have the following:*

(i) Maximisation of Φ_n over $\mathcal{C}$ is equivalent to its maximisation over $\mathcal{C}_m(\mathcal{I})$,

(ii) the maximiser

$$\widehat{\psi}_n \;:=\; \underset{\varphi\in\mathcal{C}_m(\mathcal{I})}{\arg\max}\, \Phi_n(\varphi)$$

exists and is unique.

Therefore, attention can be restricted to concave functions φ such that $\varphi = -\infty$ outside $[k_1, k_m]$ and having knot points only in the set of observations. If $\kappa_1, ..., \kappa_p$ denote the internal knot points of $\widehat{\psi}_n$, then it is not difficult to see that $\widehat{\psi}_n$ must have the following form

$$\widehat{\psi}_n(k) = a + bk + \sum_{i=1}^{p} c_i(\kappa_i - k)_+, \quad k \in \mathbb{Z} \cap [k_1, k_m] \tag{3.2}$$

where $a, b \in \mathbb{R}$ and $c_i < 0$.

Remark 3.1. *Note that given the location of the knots as in* (3.2)*, to find the MLE one needs only find the* $p+2$ *unknown values of* $a, b, c_1, \ldots, c_p$*. Note also that since we define the first and the last point of the MLE also as a knot, then the number of unknown values is exactly equal to the number of knots of the log-MLE. In addition, from Lemma 3.2, we know that the MLE satisfies* $p+1$ *equalities in* (3.4)*, plus* $\sum_x \widehat{p}_{n,x} = 1$*. Hence, we have* $p+2$ *equations with* $p+2$ *unknowns, as long as the locations of the knots are known. In essence, this tells us that the "degrees of freedom" of the estimator is equal to the number of knots. We believe that this characteristic is one key to the quality of the performance of the MLE, as compared to, for example, the empirical estimator, which has higher degrees of freedom. Further comparisons are provided in Section 5. We also make use of this heuristic when we develop our confidence bands in Section 4.4.*

3.2 Characterisation

In the study of shape-constrained estimators, characterisations provide invaluable insight into their behaviour. These are often referred to as the Fenchel conditions, due to their relationship with Fenchel duality in convex optimization problems. The characterisation of the MLE of a log-concave pmf is given below. Note that it shares a lot of similarity with the characterisation in the continuous setting (Dümbgen and Rufibach, 2009, Theorem 2.4). In what follows, $\mathbb{F}_n$ denotes the empirical cumulative distribution function.

Lemma 3.2. *Let* $\tilde{\psi} \in \mathcal{C}_m(\mathcal{I})$ *such that* $\tilde{F}_n(x) = \sum_{k=k_1}^{x} \exp \tilde{\psi}(k),\ x \in \mathbb{Z} \cap [k_1, k_m]$ *satisfies* $\tilde{F}_n(k_m) = 1$*. Then,* $\tilde{\psi} = \widehat{\psi}_n$ *if and only if the following conditions hold*

$$\sum_{j=1}^{j_x - 1} \mathbb{F}_n(k_j)(k_{j+1} - k_j) + \mathbb{F}_n(k_{j_x})(x - k_{j_x}) \geq \sum_{k=k_1}^{x-1} \tilde{F}_n(k), \quad \forall\ x \in \mathbb{Z} \cap [k_1, k_m] \tag{3.3}$$

$$= \sum_{k=k_1}^{x-1} \tilde{F}_n(k), \quad \text{if } x \text{ is a knot of } \tilde{\psi} \tag{3.4}$$

where j_x *is the unique index such that* $k_{j_x} \leq x < k_{j_x+1}$*. We use the conventions that both sums are equal to 0 if* $x = k_1$*, and that* $j_x = m$ *if* $x = k_m$*.*

The above characterisation allows us to uniquely identify the estimator in terms of simple inequalities. Furthermore, the linearity of the characterisation allows us to establish asymptotic theory for the estimator in Section 4. However, this characterisation is by no means unique, as there are other ways of uniquely identifying the MLE. Further properties of the MLE are

given in Proposition 7.1 in Appendix C. As an immediate corollary to Proposition 7.1, we obtain the following identity and bounds, where the latter hold for any $a \in \mathbb{R}$ and $m \geq 1$.

$$\sum_x x\widehat{p}_{n,x} = \sum_x x\bar{p}_{n,x} \tag{3.5}$$

$$\sum_x |x-a|^m \widehat{p}_{n,x} \leq \sum_x |x-a|^m \bar{p}_{n,x}. \tag{3.6}$$

Hence, the MLE has the same mean as the empirical distribution and a smaller variance than the empirical distribution. Similar bounds were observed in Dümbgen and Rufibach (2009) and Cule et al. (2010) for the MLE of a continuous log–concave density.

3.3 Computation of the MLE

Below, we provide only a brief description of how to compute the MLE of a log-concave pmf. For details we refer to Weyermann (2008). First, any function $\psi \in \mathcal{F}_m(\mathcal{I})$ can be identified with the vector $\boldsymbol{\psi} = (\psi(k_j))_{j=1}^m \in \mathbb{R}^m$. Conversely, each vector $\boldsymbol{\psi} \in \mathbb{R}^m$ defines the function $\psi \in \mathcal{F}_m(\mathcal{I})$ via

$$\psi(z) := \Big(1 - (z-k_j)/\delta_j\Big)\psi_j + (z-k_j)/\delta_j$$

for $z \in \mathbb{Z} \cap [k_j, k_{j+1}]$, $\delta_j = k_{j+1} - k_j$ and $1 \leq j < m$. Using this representation and using the linearity of ψ between k_j and k_{j+1} we can write the Lagrange term of l as

$$\begin{aligned}
\sum_{z \in \mathbb{Z}} \exp\psi(z) &= \left[\sum_{j=1}^{m-1}\sum_{i=0}^{\delta_j-1} \exp\Big((1-i/\delta_j)\psi_j + (i/\delta_j)\psi_{j+1}\Big)\right] + \exp\psi_m \\
&= \sum_{j=1}^{m-1}\sum_{i=0}^{\delta_j} \exp\Big((1-i/\delta_j)\psi_j + (i/\delta_j)\psi_{j+1}\Big) - \sum_{j=2}^{m-1}\exp\psi_j \\
&= \sum_{j=1}^{m-1} J_{\delta_j}(\psi_k, \psi_{k+1}) - \sum_{j=2}^{m-1}\exp\psi_j
\end{aligned}$$

where $J_{\delta_j}(\psi_j, \psi_{j+1}) = \sum_{i=0}^{\delta_j} \exp\Big((1-i/\delta_j)\psi_j + (i/\delta_j)\psi_{j+1}\Big)$, for $j \in \{1, \ldots, m-1\}$. Note the similarity of the Lagrange term to the corresponding expression in the continuous case, see Dümbgen et al. (2010, Section 2). The rewritten log-likelihood function (3.1) that we seek to maximise then amounts to

$$\Phi_n(\psi) = \sum_{j=1}^{m} w_j\psi(k_j) - \sum_{j=1}^{m-1} J_{\delta_j}(\psi_j, \psi_{j+1}) + \sum_{j=2}^{m-1}\exp\psi_j$$

which is now a concave function $\mathbb{R}^m \to \mathbb{R}$. Furthermore, a function $\psi \in \mathcal{F}_m(\mathcal{I})$ belongs to $\mathcal{C}_m(\mathcal{I})$ if and only if the corresponding vector $\boldsymbol{\psi} \in \mathbb{R}^m$ meets the following conditions:

$$\frac{\psi_j - \psi_{j-1}}{\delta_{j-1}} \geq \frac{\psi_{j+1} - \psi_j}{\delta_j}$$

for $j = 2, \ldots, m-1$ so that we end up with the task of maximizing the function $\Phi_n : \mathbb{R}^m \to \mathbb{R}$ subject to $m-2$ linear constraints. As in Dümbgen et al. (2010) an active set algorithm can then be set up to solve this maximization problem.

After streamlining it, we have translated the original Matlab code developed in Weyermann (2008) to R (R Development Core Team, 2011) and provide in the R package **logcondiscr** (Rufibach et al., 2011) the function `logConDiscrMLE` that computes the MLE from a given sample. The package also provides functions to compute the pointwise confidence bands introduced in Section 4.2 and the k-inflated log-concave pmf as discussed in Section 6.

R offers built-in functions to generate samples of random numbers from standard distributions (Binomial, negative Binomial, Geometric, Poisson). Corresponding functions for the Skellam distribution are available in the R package **skellam** (Lewis, 2009). To generate random numbers from an arbitrary log-concave pmf, the algorithm discussed in Devroye (1987) can be used. The generator is a special case of a rejection sampling algorithm and uses the fact that for any log-concave pmf with a mode at m and probabilities p_k we have

$$p_{m+k} \quad \leq \quad \min\{p_m, p_m \exp(1 - p_m|k|)\},$$

for all k. For a corresponding result for a log-concave density function see e.g. Rufibach (2006, Lemma 2.2.1).

4 Consistency and $\sqrt{n}$ asymptotics of the MLE

4.1 Consistency of the MLE

For two probability mass functions on $\mathbb{Z}$, say p and q, we define

$$\rho_{\mathrm{KL}}\left(q \,\|\, p\right) \quad = \quad \sum_{x \in \mathbb{Z}} \log\left(\frac{p_k}{q_k}\right) p_k$$

to denote the Kullback-Leibler (KL) divergence, also known as the relative entropy. Recall that $\rho_{\mathrm{KL}}\left(q \,\|\, p\right) \geq 0$, however, $\rho_{\mathrm{KL}}\left(q \,\|\, p\right) \neq \rho_{\mathrm{KL}}\left(p \,\|\, q\right)$ and hence the KL divergence is not a metric, but rather a premetric. Note that $\widehat{p}_n = \mathrm{argmin}_{q \in \mathcal{LC}_1} \rho_{\mathrm{KL}}\left(q \,\|\, \bar{p}_n\right)$, where $\bar{p}_n$ denotes the empirical pmf, and therefore the KL divergence is a natural measure to consider in the context of maximum likelihood estimation. In the following, p denotes the true pmf, and $\psi = \log p$.

Theorem 4.1. *Suppose that p is a discrete distribution on $\mathbb{Z}$ with finite mean such that $\sum_k p_k \log p_k < \infty$. Then there exists a unique log-concave pmf on $\mathbb{Z}$, $\widehat{p}$, such that*

$$\widehat{p} \quad = \quad \mathrm{argmin}_{q \in \mathcal{LC}_1} \rho_{\mathrm{KL}}\left(q \,\|\, p\right).$$

Furthermore, $\mathcal{H}(\widehat{p}_n, \widehat{p}) \to 0$ almost surely.

The proof is deferred to the Appendix. Note that by Lemma 7.3 the above convergence holds also in all other metrics described there, including pointwise convergence and ℓ_k distance for any $k \geq 1$. The result tells us that, even if the class $\mathcal{LC}_1$ was originally miss–specified, the MLE converges to the pmf which is closest, in the KL sense, to the true pmf. Of course,

if p is log-concave, then $\widehat{p} = p$, and our result implies that the MLE is consistent (note that any log–concave pmf satisfies the conditions of the theorem). Our proof follows that of Cule and Samworth (2010). It also provides an alternative way of showing existence and uniqueness of the MLE.

Corollary 4.2. *Let $\widehat{F}_n(x) = \sum_{y \le x} \widehat{p}_{n,y}$, and let $\widehat{F} = \sum_{y \le x} \widehat{p}_y$. Then*

$$\sup_{x \in \mathbb{Z}} |\widehat{F}_n(x) - \widehat{F}(x)| \quad \to \quad 0$$

almost surely.

Recall Definition 2.4 of a knot point. The following result states that knot points of the $\widehat{\psi}_n$ are also consistent.

Lemma 4.3. *For any knot point r of the logarithm of the pmf p, there exists a positive integer n_0 sufficiently large such that for all $n \ge n_0$, r is also a knot point of the MLE $\widehat{p}_n$ with probability one.*

From a theoretical point of view, Lemma 4.3 is very important in deriving weak convergence of our estimator. In practice, it implies that a knot of the logarithm of the true log–concave pmf is also a knot of the log-MLE $\widehat{\psi}_n$ when the sample size is large enough. The same lemma does not say anything about the converse property, as an observed knot of $\widehat{\psi}_n$ is not necessarily a true knot. The confidence bands derived in this work rely, however, on our knowledge of the knot points. What allows us to overcome this issue is Remark 3.1: Namely, assuming more knots than necessary only increases our degrees of freedom. In Section 4.4 we discuss further the impact of the knot points on confidence bands.

4.2 Pointwise asymptotics of the MLE

Let p denote the true log–concave pmf p and $\psi = \log p$. In what follows, we assume that p has a left-sided support. The theory in the case of a right-sided support is deduced from the latter using the transformation $x \mapsto -x$ which preserves log-concavity. Without loss of generality, we assume that the support of p is of the form $[0, \ldots, a] \cap \mathbb{N}$ with $a \in \mathbb{N} \cup \{+\infty\}$. Fix a point x which lies between two knot points $0 \le r < s$, that are a finite distance apart. This excludes distributions such as the geometric; we consider these to be the "degenerate" cases, and these will be studied elsewhere.

In the following, $\mathbb{U}$ denotes a Brownian bridge from $(0,0)$ to $(1,0)$. For $x \in \{r, \ldots, s-1\}$, define the quantities

$$\widehat{\mathbb{W}}_n(x) = \sqrt{n}(\widehat{p}_{n,x} - p_x) \qquad \mathbb{W}_n(x) = \sqrt{n}(\bar{p}_{n,x} - p_x),$$
$$\widehat{\mathbb{G}}_n(x) = \sum_{k=r}^{x} \widehat{\mathbb{W}}_n(k) \qquad \mathbb{G}_n(x) = \sum_{k=r}^{x} \mathbb{W}_n(k).$$

In addition, for $x \in \{r, \ldots, s\}$, we define

$$\widehat{\mathbb{H}}_n(x) = \sum_{k=r}^{x-1} \widehat{\mathbb{G}}_n(k) \qquad \mathbb{Y}_n(x) = \sum_{k=r}^{x-1} \mathbb{G}_n(k),$$

with the convention that $\widehat{\mathbb{H}}_n(r) = \mathbb{Y}_n(r) = 0$. It is well-known that the processes on the right, $\mathbb{W}_n, \mathbb{G}_n$, and $\mathbb{Y}_n$, have Gaussian limits. Finally, let F denote the cdf of the true pmf p. We define the least squares (LS) functional

$$\Phi(g) \quad = \quad \frac{1}{2}\sum_{x=r}^{s-1} g^2(x)p_x - \sum_{x=r}^{s-1} g(x)\mathbb{W}(x), \tag{4.1}$$

over the class of concave functions in $\mathbb{R}^{\{r,\dots,s-1\}}$, where $\mathbb{W}(x) = \mathbb{U}(F(x)) - \mathbb{U}(F(x-1))$.

Proposition 4.4. *The functional Φ in (4.1) admits a unique minimiser, g^*, over the class of concave functions on $\{r\dots,s-1\}$. Furthermore, $g = g^*$ if and only if the process $\mathbb{H}$ defined on $\{r,\dots,s\}$ satisfies*

$$\mathbb{H}(x) \begin{cases} \leq \mathbb{Y}(x), & \text{for } x \in \{r,\dots,s-1\} \\ = \mathbb{Y}(x), & \text{if } x \in \{r,\dots,s-1\} \text{ is a knot of } g^* \end{cases} \tag{4.2}$$

where

$$\mathbb{Y}(x) = \sum_{k=r}^{x-1}\sum_{j=r}^{k} \mathbb{W}(j), \quad \text{and} \quad \mathbb{H}(x) = \sum_{k=r}^{x-1}\sum_{j=r}^{k} g^*(j)p_j, \tag{4.3}$$

with the convention that $\mathbb{Y}(r) = \mathbb{H}(r) = 0$, and the boundary condition $\mathbb{Y}(s) = \mathbb{H}(s)$. Note that for $x \in \{r,\dots,s-1\}$

$$g^*(x) \quad = \quad \frac{(\Delta\mathbb{H})_x}{p_x},$$

with the further convention that $\mathbb{H}(r-1) = 0$.

We are now able to state our main asymptotic result.

Theorem 4.5. *Let $r < s$ be two successive knot points of the true pmf p, and let $\mathbb{H}$ denote the (unique) process on $\{r,\dots,s\}$ as defined in Proposition 4.4, with the additional convention that $\mathbb{H}(r-1) = 0$. Then*

$$\begin{aligned}
\sqrt{n}(\widehat{F}_n(x) - F(x)) &\overset{d}{\to} (\nabla\mathbb{H})_x + \mathbb{U}(F(r-1)), \\
\sqrt{n}(\widehat{p}_{n,x} - \widehat{p}_x) &\overset{d}{\to} (\Delta\mathbb{H})_x, \\
\sqrt{n}(\widehat{\psi}_{n,x} - \psi_x) &\overset{d}{\to} \frac{(\Delta\mathbb{H})_x}{p_x}.
\end{aligned}$$

Note that if $r = s-1$ in the above that, by definition, $(\Delta\mathbb{H})_r = \mathbb{H}(r+1) - 2\mathbb{H}(r) + \mathbb{H}(r-1) = \mathbb{H}(s) = \mathbb{W}(r)$. In other words, if the pmf p at x satisfies $(\Delta\psi)_x < 0$ and $(\Delta\psi)_{x+1} < 0$, then the limiting distribution of $\sqrt{n}(\widehat{p}_{n,x} - p_x)$ is the same as the limiting distribution of $\sqrt{n}(\bar{p}_{n,x} - p_x)$. In fact, the following stronger statement holds. Similar behaviour has been observed for the discrete Grenander estimator (Jankowski and Wellner, 2009).

Corollary 4.6. *Let x be a triple knot point of the true pmf. Then there exists an n_0 such that for all $n \geq n_0$*

$$\widehat{p}_{n,x} \quad = \quad \bar{p}_{n,x}$$

almost surely.

The situation in this discrete setting shares strong similarities with the one initially encountered by Groeneboom et al. (2001b) in convex estimation and afterwards in Balabdaoui et al. (2009) in log-concave estimation in the continuous setting. In both works, the limit distribution of the nonparametric estimators involve a stochastic process that stays above (invelope) or below (envelope) a certain Gaussian process, whose second derivative is convex (concave) and upon which depend the limit of the estimators. Knots of this second derivative are touch points of the in/en-velope and the Gaussian process.

4.3 Computation of the limiting process

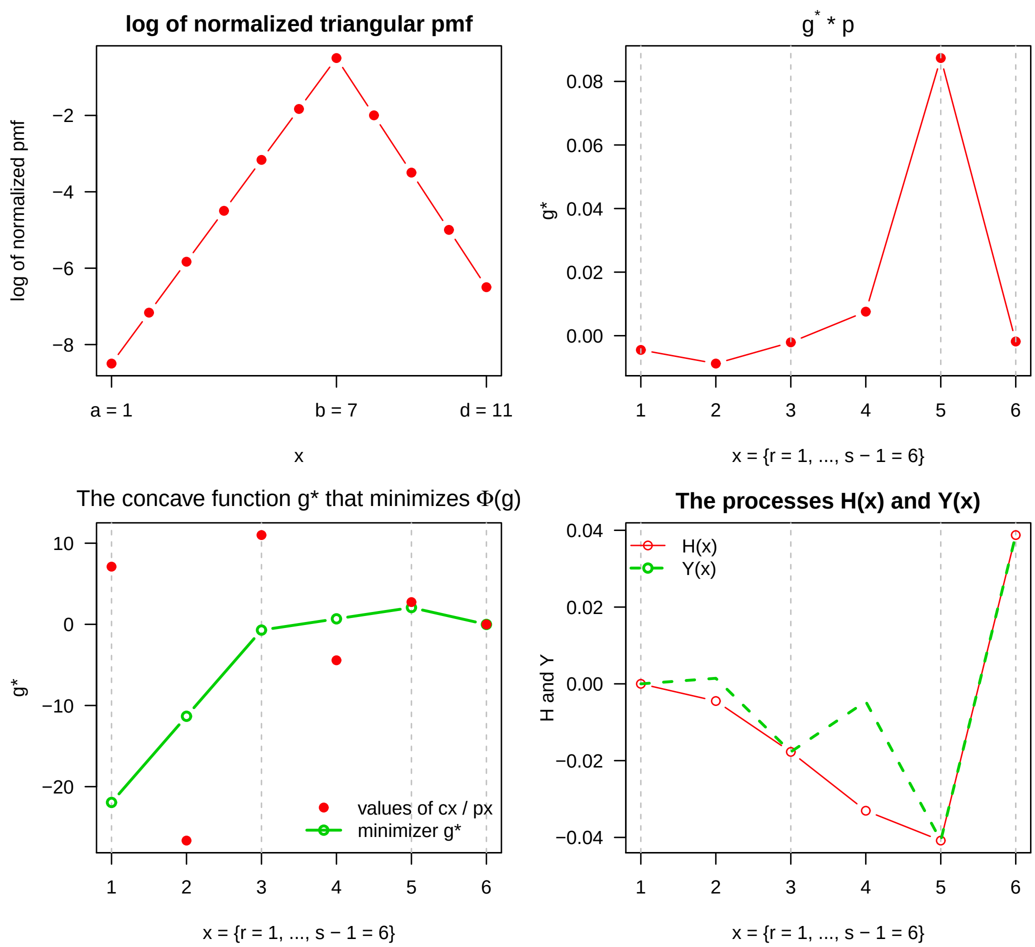

Figure 2: Triangular log-pmf, the functions g^* and $g^* p_x$ and the limit processes $\mathbb{H}$ and $\mathbb{Y}$. On the bottom left plot $cx = \mathbb{W}(x)/p_x$.

To compute $\mathbb{H}$ we need to minimise the least squares functional (4.1)

$$\Phi(g) = \frac{1}{2}\sum_{x=r}^{s-1} p_x \Big(g(x) - \frac{\mathbb{W}(x)}{p_x}\Big)^2 - \frac{1}{2}\sum_{x=r}^{s-1} \frac{\mathbb{W}(x)^2}{p_x}.$$

on the class of concave functions on $\{r, \ldots, s-1\}$, where $\mathbb{W}(x) = \mathbb{U}(F(x)) - \mathbb{U}(F(x-1))$.

Minimising Φ is then equivalent to minimising

$$C(g) = \sum_{x=r}^{s-1} p_x \left(g(x) - \frac{\mathbb{W}(x)}{p_x} \right)^2$$

over the class of concave functions on $\{r, \ldots, s-1\}$. This is a weighted least squares concave regression problem that can be numerically solved using the function `conreg` in the R package **cobs** (Ng and Maechler, 2011). Note that the vector $(\mathbb{W}(r), \ldots, \mathbb{W}(s-1))^T$ is multivariate normal with mean $\mathbf{0}$ and covariance matrix $(p_i\delta_{ij} - p_i p_j)_{i,j=1,\ldots,r-s}$.

To illustrate Proposition 4.4 we compute g^* and the processes $\mathbb{H}$ and $\mathbb{Y}$ for the triangular log-pmf supported on $\{a, \ldots, d\}$

$$p_x^{a,b,c,d,e} \quad = \quad \begin{cases} c(x-a)/(b-a) & \text{for } x \in \{a, \ldots, c\} \\ (e-c)(x-b)/(d-b) + c & \text{for } x \in \{c, \ldots, d\}, \end{cases} \tag{4.4}$$

where we choose $a = 1, b = 7, c = 8, d = 11, e = 2$ and normalise the resulting function so that the total mass is equal to one. Note that the chosen model also allows us to vary the length of a segment in a simple way by suitably defining the parameters a, b, c, d, e.

A plot of the log-pmf for the specified parameters is provided in Figure 2 (top left). The plots in Figure 2 correspond to the first segment, i.e. $r = 1$ and $s = 7$. We provide a plot (top right) of the limit of p_x as it appears in Theorem 4.5, namely $g^* \cdot p_x$ (see also Proposition 4.4), a plot of the points $\mathbb{W}(x)$ and the corresponding weighted concave regression fit as computed by `conreg` (bottom left), and finally a plot of the Gaussian process $\mathbb{Y}$ and its envelope $\mathbb{H}$ on the segment $\{r = 1, \ldots, s-1 = 6\}$.

4.4 Pointwise confidence bands

The main application of the asymptotic results described above is that they may be used to calculate pointwise confidence bands for the true pmf. For our theory to apply, we assume that the true log pmf has only finite intervals between knot points, thus excluding geometric-like distributions. However, these "degenerate" cases form only a small subset of the class of log-concave pmfs. Below, we describe how to compute 95% confidence bands, but the method can be generalised easily to any other coverage. Furthermore, we describe how to calculate the bands over the entire length of the support of the MLE. A similar approach can be also used over a smaller subsegment.

Let $\mathcal{S}$ denote the support of the true pmf, and write $\mathcal{S} = \cup_j I_j$, where $I_j = \{s_j, \ldots, s_{j+1} - 1\}$, where the s_j denote the knot points of the true log-pmf. For each $x \in I_j$, let $q_1(x), q_2(x)$ denote 2.5% and 97.5% quantiles of the distribution of $(\Delta\mathbb{H})_x$. Since we can simulate directly from the distribution of $(\Delta\mathbb{H})_x$, these are straightforward to estimate. Then,

$$\{\widehat{p}_{n,x} + q_1(x)/\sqrt{n}, \widehat{p}_{n,x} + q_2(x)/\sqrt{n}\},$$

give pointwise approximate confidence bands, which are asymptotically correct. Note that if $|I_j| = 1$, then for $x \in I_j$, $q_1(x) = -1.96\sqrt{p_x(1-p_x)}$ and $q_2(x) = +1.96\sqrt{p_x(1-p_x)}$, by Corollary 4.6.

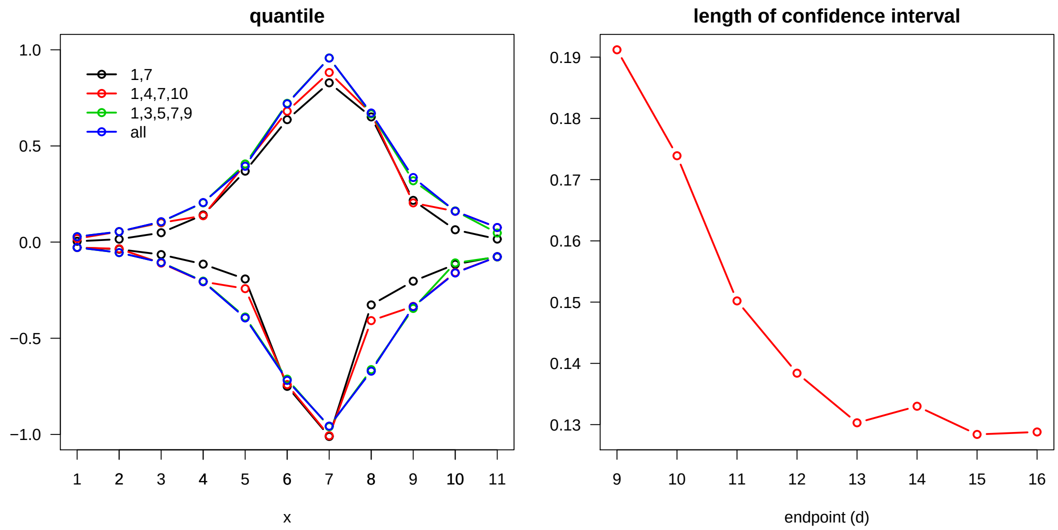

Figure 3: Estimated 95% quantiles at different points (left) and lengths of confidence intervals at one point (right) for the triangular density (4.4) with $a = 1, b = 7, c = 8, e = 2$. On the left, $d = 11$ and the quantiles were estimated from the true pmf and assuming different knot points, as indicated. On the right, the lengths of the confidence intervals were estimated using the MLE at $x = 9$ for different choices of $d \geq 9$.

To calculate the quantiles q_1 and q_2 we need to know the true pmf, including the true knot points of the log-pmf. The true pmf is easily estimated by the MLE, but a more serious issue is that we do knot know the true locations of the knot points. We propose to estimate these as the knot points of the log-MLE $\widehat{\psi}_n$. As noted following Lemma 4.3, the knot points of $\widehat{\psi}_n$ will, at worst, asymptotically overestimate the set of true knots. We believe that the penalty for this is a slight overestimation of the quantiles. Our reasoning relies on Remark 3.1 and the following discussion. Overestimating the true set of knots causes us to overestimate the degrees of freedom of the estimator, which in turn means that we have overestimated the quantiles in the confidence bands.
An example of the confidence bands for the MLE examples given in Section 1, Figure 1 for the negative binomial distribution are shown in Figure 4. The proposed confidence bands are shown in red. Confidence bands based on using all knots are shown in blue for comparison. Note that the width of the blue bands is the same as the width of pointwise bands for the empirical pmf, as they are both based on the same normal approximation. A function to compute the confidence bands is available in the R package **logcondiscr**.

5 Finite sample performance of the MLE

The results of the previous sections provide information on the performance of the MLE for very large sample sizes. To better understand the behaviour of the estimator for finite sample sizes, we have compared the results of the non-parametric vs. the parametric MLE for simulations from the Poisson ($\lambda = 2$) and negative binomial ($r = 6, p = 0.3$) distributions. In each case, we calculated

1. the empirical pmf (the MLE with no underlying assumptions),

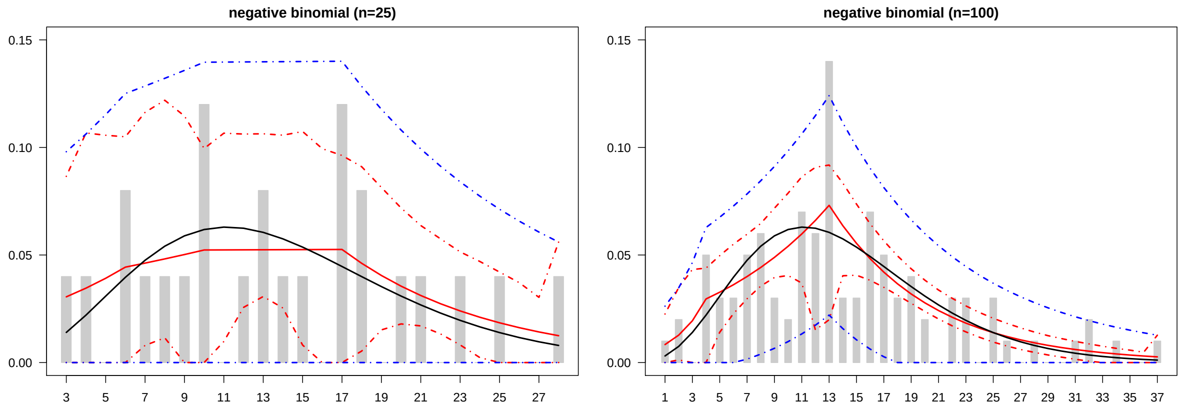

Figure 4: Nonparametric MLE of the negative binomial (6, 0.3) distribution for $n = 25$ (left) and $n = 100$ (right). The empirical distribution is shown in gray, with the MLE in red and true pmf in black. The confidence bands are shown as dashed lines, with knots based on the MLE in red, and selecting all points as knots in blue.

2. the log–concave MLE,
3. the parametric MLE assuming the geometric distribution,
4. the parametric MLE assuming the Poisson distribution,
5. the parametric MLE assuming the negative binomial distribution.

Figures 5 and 6 show boxplots of a variety of distances of each fitted distribution from the true pmf. The distances considered are the $\ell_1, \ell_2, \ell_\infty$, and the Hellinger distance, and each boxplot is the result of 1000 simulations. In Figure 5 the sample size is $n = 50$, and in Figure 6 it is $n = 500$.

The power of the non-parametric assumption is clearly shown in these simulations. The log–concave MLE performs well in estimating both distributions, albeit not as well as the correct parametric MLE. Making an incorrect parametric assumption carries with it the greatest cost, and this behaviour is much amplified when sample size is increased. Note, however, that the negative binomial MLE performs well for the Poisson distribution. This is because the negative binomial converges to the Poisson when $p = \lambda/(\lambda + r)$ and $r \to \infty$.

Lastly, the performance of the log-concave MLE is superior to that of the empirical estimator, but much more so in the negative binomial setting as compared to the Poisson setting. The asymptotic results of the previous section again provide insight into this behaviour. Figure 7 shows the pmf as well as its logarithm for both distributions. Note that the Poisson distribution exhibits more change in $\log(p)$ than the negative binomial. When the true density is strictly log–concave, Corollary 4.6 states that the empirical pmf and the log–concave MLE are asymptotically equivalent. Because the Poisson distribution is more strictly log–concave than the negative binomial, we see this behaviour much sooner in the simulations. In other words, the flatter the true $\psi = \log(p)$ is, the better the performance of the log–concave MLE over the empirical distribution. Similar behaviour was noted for the Grenander estimator of a decreasing pmf in Jankowski and Wellner (2009).

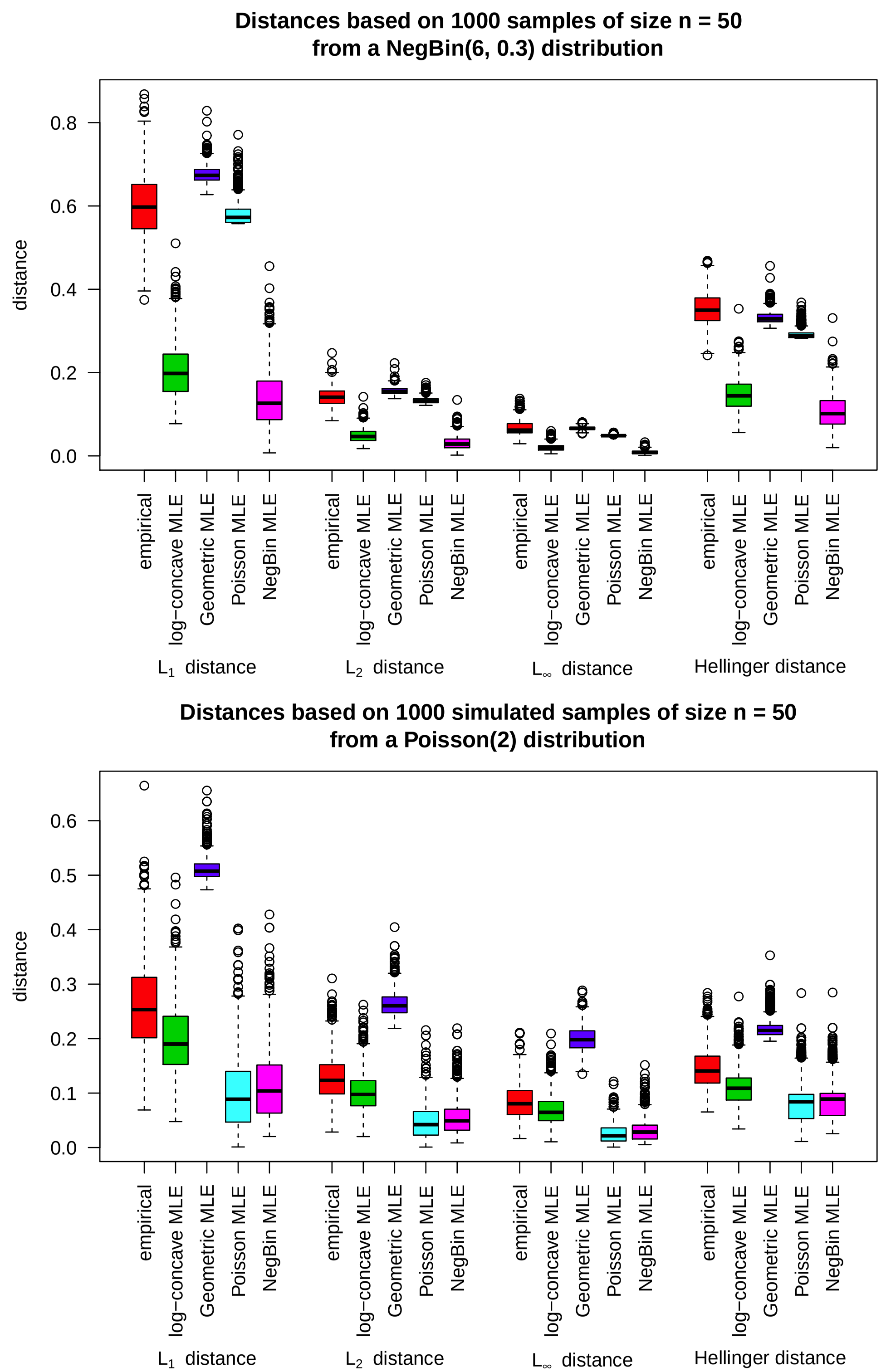

Figure 5: Assessment of performance of different estimators.

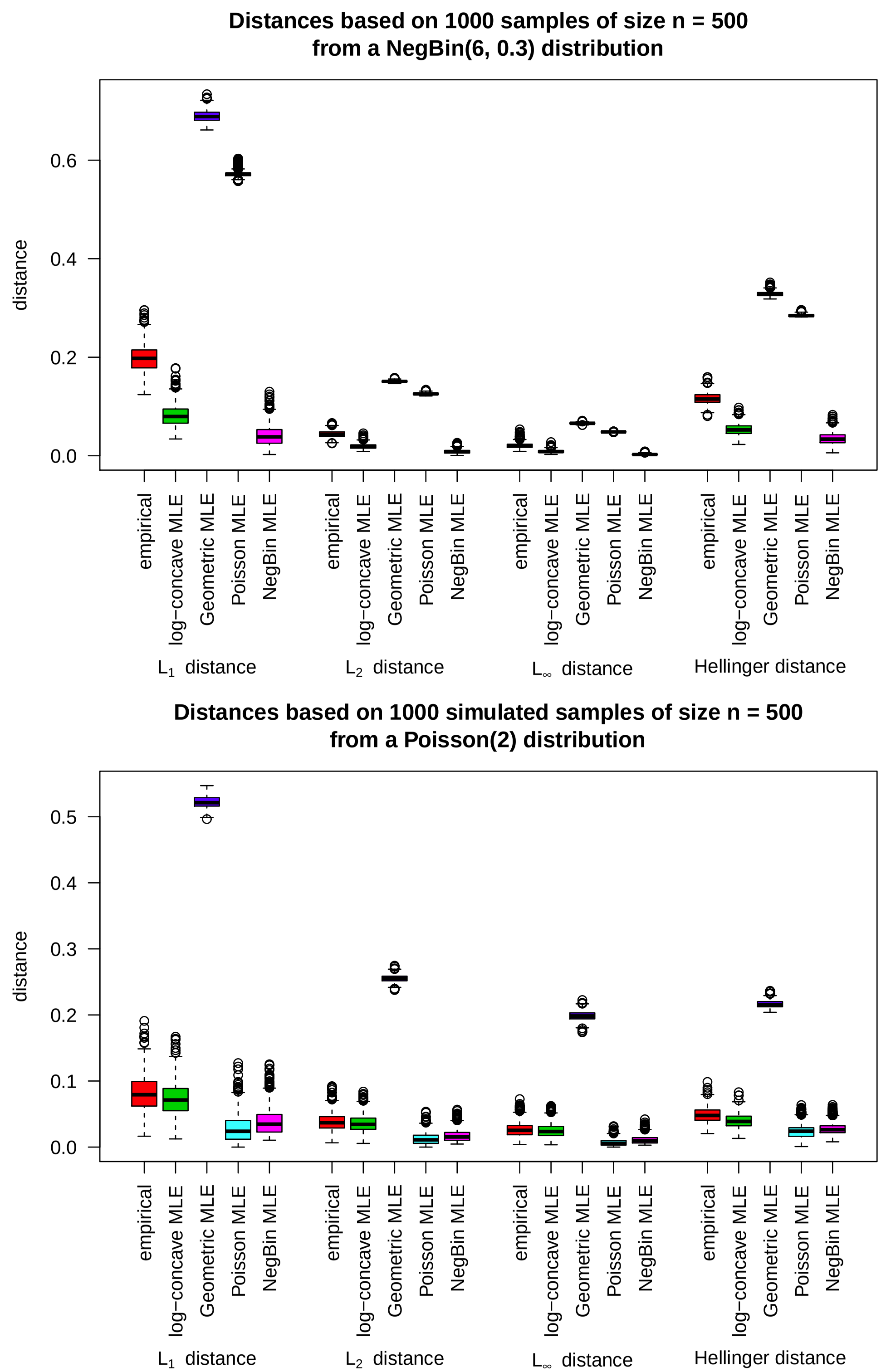

Figure 6: Assessment of performance of different estimators.

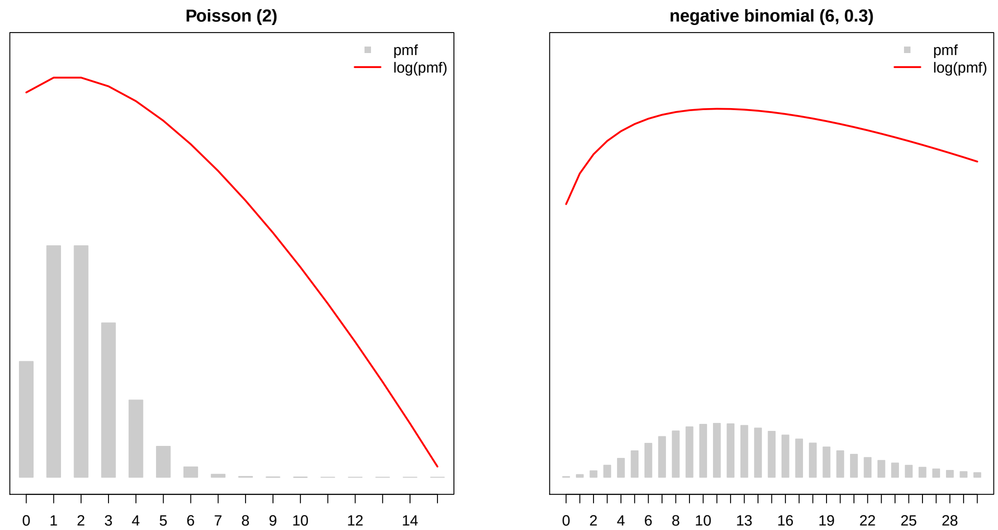

Figure 7: The probability mass function and its logarithm for the Poisson (2) and negative binomial (6, 0.3) distributions.

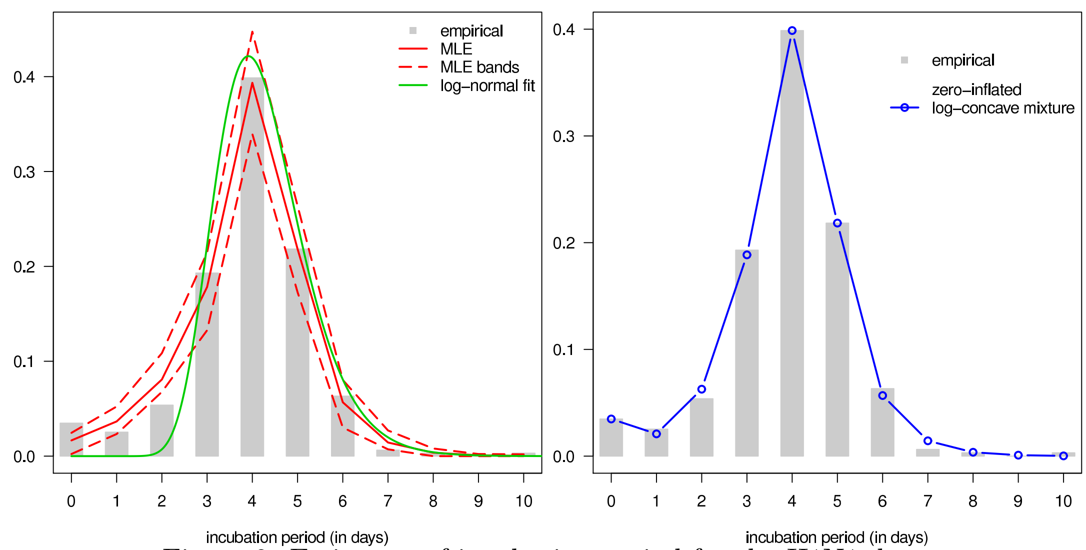

Figure 8: Estimates of incubation period for the H1N1 data.

6 Example

We illustrate the new estimator on H1N1 influenza data from Canada. Tuite et al. (2010) report an early study of the H1N1 pandemic. The goal of the study was to understand the behaviour of the disease, including the incubation period (time from exposure to the disease to onset of symptoms) and the duration of symptoms. H1N1 individual-level data was collected for laboratory–confirmed cases of the disease for a 3-month period in the spring of 2009. From these, information on the incubation period (in days, $n = 316$) and symptom duration (in days, $n = 712$) was derived. For more details on data acquisition we refer to Tuite et al.

(2010).

Clinicians and mathematical biologists are most often interested in the *fitted* mean, standard deviation, and range to understand the behaviour of the virus. These may be used in a sensitivity analysis of the developed deterministic and stochastic models, as was done in Tuite et al. (2010). For example, output of these models would be checked against known behaviour from the fitted distributions, to ascertain the appropriateness of the former. Alternatively, the model may use the fitted distribution itself within the algorithm, and hence requires the ability to simulate from the fitted distribution.

In Tuite et al. (2010), a log-normal distribution and Weibull distribution were fit to both data sets. To estimate the densities the authors used Excel's solver. After assessing goodness-of-fit, the final model chosen was the log-normal distribution. In Figure 8 (left) and Figure 9 (top), we show the log-normal fitted distributions and compare it with the log-concave MLE. Pointwise 95% confidence bands based on the MLE are also shown. It is easy to see that the log-normal does not capture well most aspects of the empirical distribution. We make the following notes about the log-concave MLE.

- Both the incubation data and duration data have been grouped, and therefore a discrete/grouped model is more appropriate than a continuous one.

- The MLE captures well the shape of the empirical distribution, including the mode and the height of the mode. Notably, the MLE has the same mean and range as the empirical distribution. As shown in (3.6), the MLE will have a smaller variance than the empirical distribution.

- Having estimated the MLE it is very easy to sample from its distribution. Given the seemingly accurate fit of our new nonparametric estimates compared to the empirical pmf, as demonstrated in Section 5, we argue that these random numbers would be more accurate than those from the log-normal model.

- Log-concavity encompasses many parametric models, but is substantially more flexible than any particular model and can capture a wide range of possible shapes. Moreover, the MLE is fully automatic, as it does not necessitate a choice of kernel, bandwidth, or prior. In this example, the MLE fits the empirical well, and it also "smooths" the empirical especially in the rather variable tail of the symptom duration distribution.

- In the analysis of an infectious disease, the incubation period is of great importance, particularly so in the lower tail of the distribution, as this provides information on the rate of spread of the virus within a population. The log-normal does not fit the lower tail of the empirical distribution, and the MLE is better at describing this behaviour. A closer examination of the empirical data shows a spike at zero, which is most likely caused by inaccurate reporting of the onset of symptoms. To better describe this behaviour, we also fit a mixture of a log-concave pmf with a point mass at zero. This is, essentially, a zero-inflated log-concave distribution. To perform the fitting procedure, we used the EM algorithm. The results are shown in Figure 8 (right). The mean of the pure MLE was 3.88, which is equal to the mean of the data. The mean of the log–concave part of the mixture model was slightly higher, at 4.02.

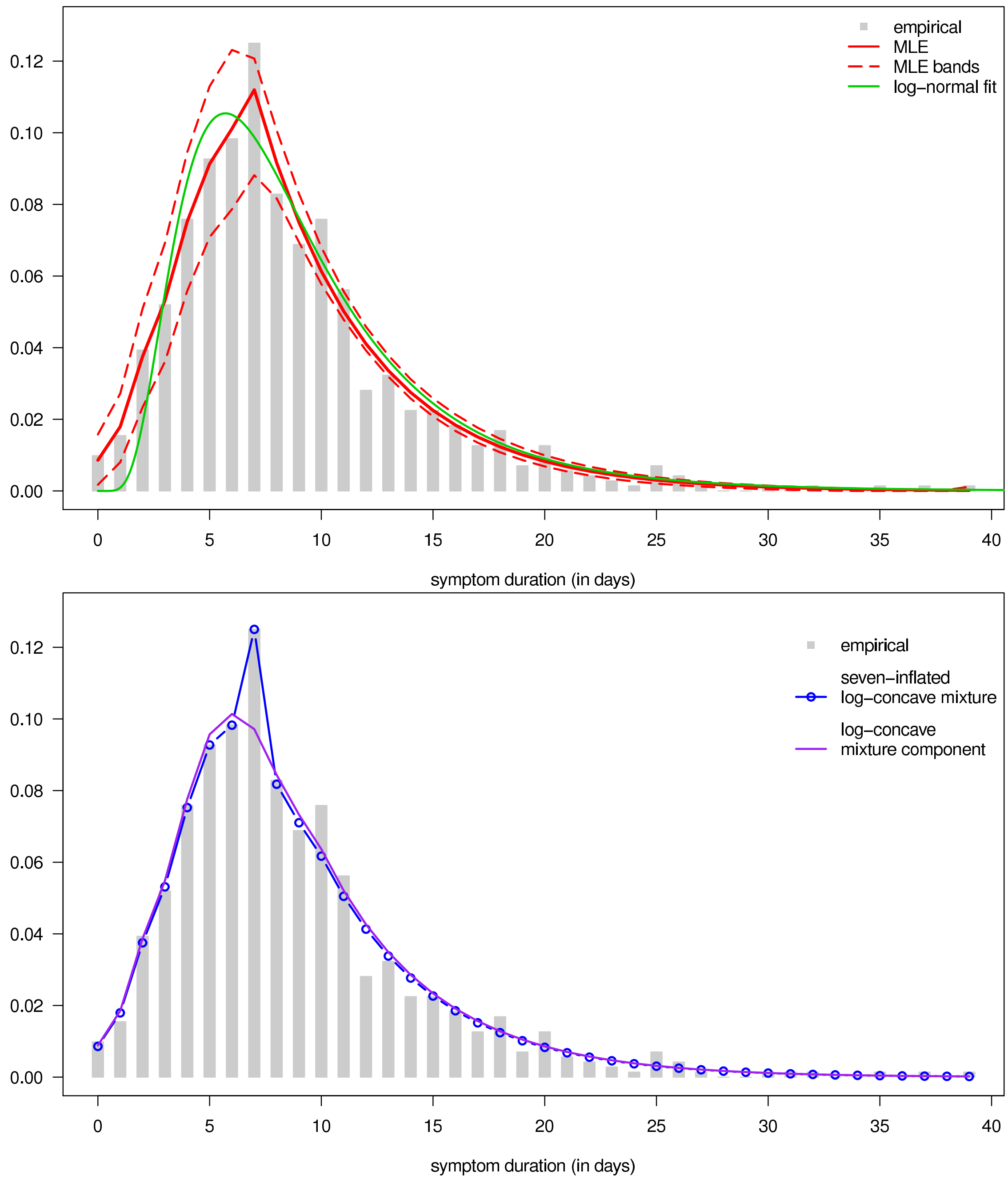

Figure 9: Estimates of duration for H1N1 data.

- The data for the duration of symptoms of the swine flu has a clear spike at $t = 7$ days. As above, this is probably caused by mis-reporting, as seven days is equivalent to one week, and therefore a likely choice in a patient's response. One ad-hoc method to account for this, is to again fit an inflated model, this time placing the mass function at $t = 7$ days. The results are shown in Figure 9 (bottom). The mean of the pure MLE was 8.66, which was also the mean of the fitted mixture model. The mean of the log–

concave part of the mixture model is slightly higher, at 8.72, however, the log-concave component has a lower mode at $t = 6$. The probability of observing an "inflated" value at seven was found to be 0.031.

- In addition to the aforementioned issue, it is quite likely that the duration data collected suffers from length-bias (see e.g. Asgharian and Wolfson, 2002), in that those with longer duration of symptoms were more likely to be observed. It would be of interest to see if our methods can be modified to include a length-bias correction, however, this is beyond the scope of this work.

7 Discussion

Fitting discrete data that exhibit unimodality is one motivation of using the log–concave MLE. Assuming that such data are drawn from a log–concave distribution is an attractive solution in the absence of an explicit characterization of a nonparametric unimodal estimator or a good justification of some parametric model. As was proved in Section 4, the log–concave MLE will converge almost surely to the true distribution when this distribution is in fact log-concave. Furthermore, the estimator will still converge to something meaningful in case the assumption of log-concavity does not hold. These results parallel those of Cule and Samworth (2010) for the log-concave MLE on $\mathbb{R}^d$.

In Section 4.2, we establish pointwise asymptotic theory for the MLE. The limit is described via a process that acts as an envelope of a finite dimensional Gaussian process on any given *finite* segment on which the true pmf is log-linear. The assumption that the log-pmf admits knots which are a finite distance apart is crucial for the derivation of the asymptotic distribution. Hence, the geometric and double-geometric distributions are excluded, as is any part of a log-concave distribution which is log-linear on an infinite set of integers. Showing existence of the envelope process involved in the limit of the log–concave MLE (and hence existence of this limit) relies on showing that the solution of a well posed least squares problem is fully characterised by the properties of this envelope. As already noted above, our approach is much inspired by the pioneering work of Groeneboom et al. (2001a) and Groeneboom et al. (2001b) in convex estimation, and also by the work of Balabdaoui et al. (2009) in log-concave estimation for absolutely continuous distributions. The main idea, which had to be re-adapted to the discrete setting, is that the limit inherits the same (characterizing) properties of its finite sample counterpart. It is the "linearisation" of the shape constraint which transforms log-concavity to concavity that naturally gives rise to the relevant least squares problem. In a sense, maximization of the log-likelihood for a finite sample has been transformed into minimization of a least squares criterion for the "infinite" sample. It is of much interest to note that some features are intrinsically common to the asymptotic theory developed here and in Groeneboom et al. (2001b) and Balabdaoui et al. (2009). Others are evidently more specific to each setting. This is summarised in Table 1 where we compare the results for log-concave estimation. We use the usual notation $\mathbb{H}$ and $\mathbb{Y}$ for the envelope and Gaussian process, and ψ for the logarithm of the density in both discrete and continuous settings.

It is well-known that in the context of shape-constrained estimation, the asymptotic behaviour is different in the "strict" and in the "degenerate" case. For example, a decreasing density satisfies $f'(x) \leq 0$. For the Grenander estimator of a decreasing density, if the true density

satisfies $f'(x_0) < 0$, then we observe an $n^{1/3}$ rate of convergence (Groeneboom, 1985). This is the “strict” case. The degenerate case occurs when $f'(x) = 0$ over some region, and here the convergence is of rate $\sqrt{n}$ (Carolan and Dykstra, 1999). Furthermore, the limiting distribution is different in these two cases. For other shape-constraints on $\mathbb{R}$, only the asymptotic behaviour in the strict setting has been established, to our knowledge. In particular, for estimation of a log-concave density on $\mathbb{R}$, Balabdaoui et al. (2009) showed that if $\psi(x) = \log f(x)$ satisfies $\psi''(x_0) < 0$, then pointwise convergence occurs at a rate of $n^{2/5}$. Asymptotic behaviour when $\psi''(x) = 0$ over some region is still an open problem.

	continuous	discrete	
	$\psi''(x) < 0$	$(\Delta\psi)_x < 0$	$(\Delta\psi)_x = 0$
Rate of convergence	$n^{2/5}$	$n^{1/2}$	
$\mathbb{H}$	$\mathbb{H}''$ is concave	$(\Delta\mathbb{H})_x/p_x$ is concave	
$\mathbb{Y}$	involves integral of Brownian motion	involves cumulative sum of Brownian bridge	
Limit of the MLE at x	involves $\mathbb{H}''(0)$	$(\Delta\mathbb{H})_x = \mathbb{W}(x)$	$(\Delta\mathbb{H})_x$

Table 1: Comparison of components of limiting theory for log-concave estimation of a probability density and a pmf.

Our results establish limiting distributions in the strict ($(\Delta\psi)_x < 0$) and degenerate ($(\Delta\psi)_x = 0$) cases for a finite region. As such, they give insight into the asymptotic behaviour in the degenerate case in the continuous setting, and we conjecture that the limiting distribution will be characterised by a solution to the LS problem

$$\text{argmin}_{g \text{ concave on } (a,b)} \left\{ \frac{1}{2} \int_a^b f_0(x) g^2(x) - \int_a^b g(x) d\,\mathbb{U}(F(x)) \right\},$$

where f_0 is the true log-concave density, $\mathbb{U}$ denotes again the Brownian bridge, and (a, b) is the largest interval where $\psi'' = 0$.

We also note that our results are similar to those of Jankowski and Wellner (2009) where the Grenander estimator for the discrete decreasing pmf was studied. There it was shown that when $(\nabla p)_x < 0$ over a region, then the MLE is asymptotically equivalent to the empirical pmf, and if $(\nabla p)_x = 0$ over a region, then the MLE's limit is described in terms of a LS problem. For the decreasing pmf there does not exist a pmf with $(\nabla p)_x = 0$ over an infinite region, and therefore the infinite region problem is unique to the log-concave setting.

For the Grenander estimator, Jankowski and Wellner (2009) also describe global convergence rates for the MLE estimator. However, the problem there is considerably simpler than for the log-concave MLE. Without going into too much detail, for the Grenander estimator the limit is described in terms of the underlying process $\mathbb{W}(x) = \mathbb{U}(F(x + 1)) - \mathbb{U}(F(x))$, which is well-defined. On the other hand, the limit for the log-concave MLE is described in terms of the underlying process $\mathbb{W}(x)/\sqrt{p_x}$, which is not well-defined globally (see e.g. Varadhan, 1968). Therefore, it appears that more subtle technical tools will need to be developed to study global convergence in this setting. We intend to explore this problem in a future work.

Acknowledgments

We thank Marios Pavlides for having shared with us some references on the subject and thoughts on the definition of log–concavity. We would like to thank Filippo Santambrogio and Jon Wellner for helpful discussions and Kathrin Weyermann and Lutz Dümbgen for sending us a copy of Kathrin's Master's thesis and the Matlab code to compute the MLE. Finally, we would like to thank Ashleigh Tuite and David Fisman who made the H1N1 data available to us, and Jane Heffernan who provided us with invaluable insight into the data.

Appendix A

Throughout the paper we use the following notation:

- the words *sequence* (indexed by $\mathbb{Z}$) and *function* (defined on $\mathbb{Z}$) interchangeably. In the first case, the k-th term of a sequence p will be denoted by p_k and the whole sequence by $\{p_k, k \in \mathbb{Z}\}$. In the second case, the value of the function $p : \mathbb{Z} \to [-\infty, \infty)$ at $k \in \mathbb{Z}$ will be denoted by $p(k)$,
- $\mathcal{C}$ to denote the class of all concave functions $\psi : \mathbb{Z} \to [-\infty, \infty)$,
- $\mathcal{LC}_1$ to denote the class of all log-concave probability mass functions,
- $\mathcal{LC}$ to denote the cone of all log-concave, non-negative sequences indexed by $\mathbb{Z}$; i.e. the set of all non-negative sequences p such that $p_k \geq 0, k \in \mathbb{Z}$, and satisfying the properties A and B of Definition 2.3 below,
- $z_+ = 1_{z \geq 0}$,
- $\delta_{ij} = 1$ if $i = j$, and 0 otherwise,
- $\ell_k(p, q)$ for the distance $\left(\sum_{k \in \mathbb{Z}} (p_k - q_k)^k\right)^{1/k}$ if $1 \leq k < \infty$, and $\sup_{k \in \mathbb{Z}} |p_k - q_k|$ if $k = \infty$ for two probability mass functions $\{p_k, k \in \mathbb{Z}\}$ and $\{q_k, k \in \mathbb{Z}\}$,
- $\mathcal{H}$ for the Hellinger distance, that is, $\mathcal{H}^2(p, q) = 2^{-1} \sum_{k \in \mathbb{Z}} (\sqrt{p_k} - \sqrt{q_k})^2$,
- For a sequence $\{\varphi_x, x \in \mathbb{Z}\}$, $(\nabla\varphi)_x = \varphi_{x+1} - \varphi_x$ denotes the discrete gradient, and $(\Delta\varphi)_x = \varphi_{x+1} - 2\varphi_x + \varphi_{x-1}$ denotes the discrete Laplacian.

Appendix B

In this section, we provide a proper definition of discrete log-concavity and provide a discussion of some parametric distributions that are log-concave.

7.1 Discrete concavity and log–concavity

Consider a function $f : \mathbb{Z} \to [-\infty, \infty)$ such that $f \neq -\infty$. We will say that f is concave if the piecewise linear function $\tilde{f} : \mathbb{R} \to [-\infty, \infty)$ defined as

$$\tilde{f}(x) = (k + 1 - x)f(k) + (x - k)f(k + 1), \quad \text{for } x \in [k, k + 1] \text{ and for all } k \text{ in the domain}$$

is concave, in the sense that

$$\mathrm{sub}(\tilde{f}) \quad = \quad \{(x,\mu) \in \mathbb{R}^2 : \tilde{f}(x) \geq \mu\},$$

the subgraph of $\tilde{f}$, is a convex set in $\mathbb{R}^2$; see Rockafellar (1970). Let $\mathrm{dom}(\tilde{f})$ denote the effective domain of $\tilde{f}$, that is,

$$\mathrm{dom}(\tilde{f}) \quad = \quad \{x \in \mathbb{R} : \ \exists \mu \in \mathbb{R} \text{ such that } (x,\mu) \in \mathrm{sub}(\tilde{f})\} = \{x \in \mathbb{R} : \ \tilde{f}(x) > -\infty\}.$$

Then it is easily seen that $\mathrm{dom}(\tilde{f})$ is convex, and hence an interval in $\mathbb{R}$. Also, concavity of $\tilde{f}$ is equivalent to concavity of its restriction on $\mathrm{dom}(\tilde{f})$. Note that excluding the value ∞ from the set of values of f avoids us having to deal with the undefined expression $\infty - \infty$. Also, this makes $\tilde{f}$ a finite concave function on $\mathrm{dom}(\tilde{f})$, taking possibly $-\infty$ outside this domain.
Now, if we define the effective domain of the original function f to be

$$\mathrm{dom}(f) = \{k \in \mathbb{Z} : f(k) > -\infty\}$$

then concavity of f implies that $\mathrm{dom}(f)$ is a convex set in $\mathbb{Z}$. Indeed, we can write

$$\mathrm{dom}(\tilde{f}) = \bigcup_{k \in \mathbb{Z}} \Big\{[k, k+1] \text{ such that } k \text{ and } k+1 \in \mathrm{dom}(f)\Big\}.$$

It follows that $\mathrm{dom}(\tilde{f})$ is an interval in $\mathbb{R}$ if and only if $\mathrm{dom}(f)$ is a set of consecutive integers; i.e., is convex. Finally, the function f is concave on $\mathbb{Z}$ if and only if its restriction on $\mathrm{dom}(f)$ is concave. This follows easily from noting that concavity of the restriction of f on $\mathrm{dom}(f)$ is equivalent to concavity of $\tilde{f}$ on $\mathrm{dom}(\tilde{f})$.
Hence, to construct a concave function f on $\mathbb{Z}$, it is necessary and sufficient to consider a function that is finite and concave on some nonempty subset $\mathcal{S}$ of consecutive integers and extend it to a concave function on $\mathbb{Z}$, in case $\mathcal{S} \neq \mathbb{Z}$, by setting $f = -\infty$ outside $\mathcal{S}$.

It follows that given a convex subset $\mathcal{S} := \{k_1, ..., k_2\}$ of $\mathbb{Z}$, a function $f : \mathbb{Z} \to [-\infty, \infty)$ such that f is finite on $\mathcal{S}$ and $f = -\infty$ on $\mathbb{Z} \setminus \mathcal{S}$ is concave if and only if the left and right derivatives of $\tilde{f}$ on the real interval (k_1, k_2) are nonincreasing. This is in turn equivalent to having nonincreasing order of these derivatives for two adjacent sub-intervals $[k-1, k]$ and $[k, k+1]$; i.e.,

$$f(k+1) - f(k) \leq f(k) - f(k-1), \ k \in \{k_1+1, ..., k_2-1\}$$

or equivalently

$$f(k) \geq \frac{f(k-1) + f(k+1)}{2}, \ k \in \mathbb{Z} \tag{7.1}$$

using the conventions $-\infty + c = -\infty$ for $c \in \mathbb{R} \cup \{-\infty\}$ and $-\infty \geq -\infty$. Property 7.1 is usually referred to as *mid-concavity.*
In conclusion, a function $f : \mathbb{Z} \to [-\infty, \infty)$ such that $f \neq -\infty$ is concave if and only if there exists a nonempty set $\mathcal{S}$ of $\mathbb{Z}$ of consecutive integers such that

1. f is finite on $\mathcal{S}$,

2. $f = -\infty$ on $\mathbb{Z} \setminus \mathcal{S}$,

3. f is mid-concave on $\mathbb{Z}$; i.e., satisfies the property in (7.1).

The definition of concavity of a sequence $p = \{p_k, k \in \mathbb{Z}\}$ is straightforward and follows from the identification of the terms p_k with the value at k of a function defined on $\mathbb{Z}$.

We can give a natural definition of log-concavity of a sequence $p = \{p_k, k \in \mathbb{Z}\}$ such that $p < \infty$. Such a sequence p is said to be log-concave if $\log p = \{\log p_k, k \in \mathbb{Z}\}$ is concave, that is, there exists a nonempty set $\mathcal{S}$ of consecutive integers such that

1. $\log p_k > -\infty$, for $k \in \mathcal{S}$,

2. $\log p_k = -\infty$, for $k \in \mathbb{Z} \setminus \mathcal{S}$,

3. $\log p_k \geq 1/2\ (\log p_{k-1} + \log p_{k+1})$, for all $k \in \mathbb{Z}$.

This corresponds exactly to Definition 2.3 whose Property A is equivalent to 1 and 2, and Property B to 3.

7.2 Some examples of log–concave pmf's

Let $n, r \in \mathbb{N}^*$, $\alpha \in (0,1)$ and $\theta, \mu \in (0, \infty)$. The following parametric models are some examples of discrete distributions admitting a log–concave pmf.

1. A binomial distribution with pmf

$$p_k \quad = \quad \begin{cases} \binom{n}{k} \alpha^k (1-\alpha)^{n-k}, & k \in \{0, ..., n\} \\ 0, & \text{otherwise} \end{cases}$$

is log-concave with bounded support $\mathcal{S} = \{0, ..., n\}$ and a ratio sequence $\{p_k/p_{k-1}, k \in \{0, ..., n\}\} = \{\infty, n\alpha(1-\alpha)^{-1}, ..., (n-k+1)k^{-1}\alpha(1-\alpha)^{-1}, ..., n^{-1}\alpha(1-\alpha)^{-1}\}$. The mode of p is located at the smallest integer $k \in \{0, ..., n\}$ such that $(n-k+1)k^{-1}\alpha(1-\alpha)^{-1} \geq 1$ and $(n-k)(k+1)^{-1}\alpha(1-\alpha)^{-1} \leq 1$, that is, $k \in [(n+1)\alpha - 1, (n+1)\alpha]$.

2. A negative binomial with pmf

$$p_k \quad = \quad \begin{cases} \binom{k+r-1}{k} (1-\alpha)^r \alpha^k, & k \in \mathbb{N} \\ 0, & \text{otherwise} \end{cases}$$

is log–concave with unbounded support $\mathcal{S} = \mathbb{N}$ and a ratio sequence $\{p_k/p_{k-1}, k \in \mathbb{N}\} = \{\infty, \alpha r, \ldots, \alpha(k+r-1)/k, \ldots\}$. he mode of p is located at the smallest integer $k \in \{0, ..., n\}$ such that $\alpha(k+r-1)/k \geq 1$ and $\alpha(k+r)/(k+1) \leq 1$, that is, $k \in [(\alpha r - 1)/(1-\alpha), (\alpha r - 1)/(1-\alpha) + 1]$.

3. A geometric distribution with pmf

$$p_k \quad = \quad \begin{cases} \alpha(1-\alpha)^k, & k \in \mathbb{N} \\ 0, & \text{otherwise} \end{cases}$$

is log-concave with unbounded support (from one side) $\mathcal{S} = \mathbb{N}$ and, except for the first term, a constant ratio sequence $\{p_k/p_{k-1}, k \in \mathbb{N}\} = \{\infty, 1-\alpha, ..., 1-\alpha, ...\}$. The mode of p is located at 0.

4. A Poisson distribution with pmf

$$p_k = \begin{cases} e^{-\theta}\theta^k/k!, & k \in \mathbb{N} \\ 0, & \text{otherwise} \end{cases}$$

is log-concave with unbounded support (from one side) $\mathcal{S} = \mathbb{N}$ and a ratio sequence $\{p_k/p_{k-1}, k \in \mathbb{N}\} = \{\infty, \theta, \theta/2, ..., \theta/k, ...\}$. The mode is located at the smallest integer $k \in \mathbb{N}$ such that $k \in [\theta - 1, \theta]$.

5. A Skellam distribution with pmf

$$p_k = e^{-(\theta+\mu)} \sum_{m=0}^{\infty} \frac{\theta^{k+m}\mu^m}{m!(k+m)!}, \ k \in \mathbb{Z}$$

is log-concave with unbounded support $\mathcal{S} = \mathbb{Z}$. Log-concavity of the Skellam distribution is a consequence of the fact that it is the distribution of the difference of two independent Poisson random variables, here with parameters θ and μ. Indeed, it is not hard to see that if X is a log-concave random variable, then so is $-X$. The claim of log-concavity follows by appealing to closedness of the class of log-concave distribution under convolution (see Theorem 2 of Keilson and Gerber, 1971). For $k \in \mathbb{Z}$, we have

$$\frac{p_k}{p_{k-1}} = \sqrt{\frac{\theta}{\mu}} \frac{I_{|k|}(\sqrt{\theta\mu})}{I_{|k-1|}(\sqrt{\theta\mu})}$$

where I_m is the modified Bessel function of the first kind and order m (see e.g. Abramowitz and Stegun, 1964, Section 9.6). The Skellam distribution was originally introduced by Skellam (1946) and found applications in many fields, such as modeling the difference of the number of goals in football games, and the intensity difference of pixels in cameras. For more details about this distribution, we refer the reader for example to Karlis and Ntzoufras (2003), Karlis and Ntzoufras (2005), Karlis and Ntzoufras (2006), Karlis and Ntzoufras (2009), and Alzaid and Omair (2010) and the references therein.

Appendix C

This section contains all proofs and relevant technical details.

7.3 Proof from Section 2

Proof of Proposition 2.2. The result follows immediately from Dharmadhikari and Joag-Dev (1988, Theorem 2.8) since then

$$P(X \in A_i)^2 \geq P(X \in A_{i+1})P(X \in A_{i-1}),$$

from which it follows that $\{p_i\}$ is log-concave. □

7.4 Proofs from Section 3.1

Proof of Theorem 3.1. We give first a proof of (i). Let φ be a concave function in $\mathcal{C}$ such that $\Phi_n(\varphi) > -\infty$. Let now ψ be the unique concave function in $\mathcal{C}_m(\mathcal{I})$ such that $\psi(k_i) = \varphi(k_i)$ for $i \in \{1, ..., m\}$ and interpolating between the observations, that is if $k_i < k_{i+1}$ for some $i \in \{1, ..., m-1\}$, then for any $k \in \{k_i + 1, ..., k_{i+1} - 1\}$ we have

$$\psi(k) = \psi(k_i) + \frac{\psi(k_{i+1}) - \psi(k_i)}{k_{i+1} - k_i}(k - k_i).$$

Since φ is concave, we have the following properties:

1. $\psi \le \varphi$ on $\mathbb{Z} \cap [k_1, k_m]$,
2. $\Phi_n(\psi) \ge \Phi_n(\varphi)$ with equality if the two functions coincide.

This implies that in maximizing the criterion Φ_n, the latter does not decrease when we replace a function in $\mathcal{C}$ by a candidate in the class $\mathcal{C}_m(\mathcal{I})$. Hence, we can restrict attention to functions in the latter class which formally means that if a maximiser of Φ_n exists, we should look for it in the smaller class $\mathcal{C}_m(\mathcal{I})$.

Next, we prove (ii). Let $(\psi^p)_p$ be a maximizing sequence of the criterion Φ_n in $\mathcal{C}_m(\mathcal{I})$; i.e., $\lim_{p\to\infty} \Phi_n(\psi^p) = \sup_{\psi \in \mathcal{C}_m(\mathcal{I})} \Phi_n(\psi)$. We will show now that the maximization problem can be restricted to a compact subset of $\mathcal{C}_m(\mathcal{I})$.
Consider the functions $f_j(x) = w_j x - \exp(x), j = 1, \ldots, m$. It is easy to see that

$$\Phi_n(\psi^p) \le \sum_{j=1}^{m} f_j(\psi^p(k_j)). \tag{7.2}$$

Now, suppose that there exists $j \in \{1, \ldots, m\}$ such that $|\psi^p(k_j)| \to \infty$ as $p \to \infty$. From (7.2) and coercivity of f_j, it follows that $\Phi_n(\psi^p) \to -\infty$ which yields a contradiction since $(\psi^p)_p$ is a maximizing sequence of Φ_n (we should have at least $\lim_{p\to\infty} \Phi(\psi^p) \ge -(k_m - k_1 + 1)$, the value taken by the criterion for the sequence whose all terms at equal to zero.) Hence, there exists $K > 0$ such that for all $j \in \{1, \ldots, m\}$, $|\psi^p(k_j)| \le K$. Since ψ^p is linear between the observations $k_j, j = 1, \ldots, m$, it is easy to see that this bound actually applies for all $k \in \{k_1, k_1 + 1, \ldots, k_m - 1, k_m\} = \mathbb{Z} \cap [k_1, k_m]$. It follows that the closed subset (concavity is defined by a set of non-strict inequalities)

$$\Big\{\psi : \psi \in \mathcal{C}_m(\mathcal{I}) \text{ and } \max_{k \in \mathbb{Z} \cap [k_1, k_m]} |\psi(k)| \le K\Big\}$$

is bounded, and hence compact. Continuity of Φ_n concludes the proof of existence. Uniqueness follows from strict concavity of Φ_n. $\square$

7.5 Proofs from Section 3.2

In the sequel, we will make extensive use of suitable perturbation functions; i.e., functions which determine an admissible direction in which the MLE is perturbed. All these results will be based on the following proposition, which gives suitable directional derivatives of Φ_n.

Proposition 7.1. *Let $\epsilon_0 > 0$. Consider functions P_1 and P_2 defined on $\mathbb{Z}$ such that $\widehat{\psi}_n + \epsilon P_1 \in \mathcal{C}$ for all $\epsilon \in (0, \epsilon_0)$ and $\widehat{\psi}_n + \epsilon P_2 \in \mathcal{C}$ for all $\epsilon \in (-\epsilon_0, \epsilon_0)$. Then,*

$$\sum_{j=1}^{m} \left(\mathbb{F}_n(k_j) - \mathbb{F}_n(k_{j-1})\right) P_1(k_j) \quad \leq \quad \sum_{k=k_1}^{k_m} \left(\widehat{F}_n(k) - \widehat{F}_n(k-1)\right) P_1(k) \tag{7.3}$$

and

$$\sum_{j=1}^{m} \left(\mathbb{F}_n(k_j) - \mathbb{F}_n(k_{j-1})\right) P_2(k_j) = \sum_{k=k_1}^{k_m} \left(\widehat{F}_n(k) - \widehat{F}_n(k-1)\right) P_2(k). \tag{7.4}$$

Proof. Using the convention $\mathbb{F}_n(k_0) = \widehat{F}_n(k_0) = 0$ and the fact that $\widehat{\psi}_n$ is the maximiser of Φ_n, we can write

$$\begin{aligned} 0 \quad &\geq \quad \lim_{\epsilon \searrow 0} \frac{\Phi_n(\widehat{\psi}_n + \epsilon P_1) - \Phi_n(\widehat{\psi}_n)}{\epsilon} \\ &= \quad \sum_{j=1}^{m} w_j P_1(k_j) - \sum_{k=k_1}^{k_m} \exp \widehat{\psi}_n(k) P_1(k) \\ &= \quad \sum_{j=1}^{m} \left(\mathbb{F}_n(k_j) - \mathbb{F}_n(k_{j-1})\right) P_1(k_j) - \sum_{k=k_1}^{k_m} \left(\widehat{F}_n(k) - \widehat{F}_n(k-1)\right) P_1(k) \end{aligned}$$

where we take $k_0 = k_1 - 1$. This yields (7.3). To get (7.4) we use the fact that the limit above is equal to zero when replacing P_1 by P_2. □

Proof of Lemma 3.2. First, we assume that $\tilde{\psi} = \widehat{\psi}$. For $x \in \mathbb{Z} \cap [k_1, k_m]$, consider the perturbation function

$$P_x(k) \quad = \quad -(x-k)_+. \tag{7.5}$$

Then, for all $t \geq 0$ the function $\tilde{\psi} + tP_x$ is concave. By Proposition 7.1, we have that

$$0 \quad \geq \quad -\sum_{j=1}^{m} \left(\mathbb{F}_n(k_j) - \mathbb{F}_n(k_{j-1})\right)(x-k_j)_+ + \sum_{k=k_1}^{k_m} \left(\tilde{F}_n(k) - \tilde{F}_n(k-1)\right)(x-k)_+.$$

It follows that

$$\begin{aligned} 0 \quad &\geq \quad -\mathbb{F}_n(k_m)(x-k_m)_+ - \sum_{j=1}^{m-1} \mathbb{F}_n(k_j)\left((x-k_j)_+ - (x-k_{j+1})_+\right) \\ &\qquad + \tilde{F}_n(k_m)(x-k_m)_+ + \sum_{k=k_1}^{k_m-1} \tilde{F}_n(k)\left((x-k)_+ - (x-(k+1))_+\right) \\ &= \quad -\sum_{j=1}^{j_x-1} \mathbb{F}_n(k_j)(k_{j+1}-k_j) + \mathbb{F}_n(k_{j_x})(x-k_{j_x}) + \sum_{k=k_1}^{x-1} \tilde{F}_n(k) \end{aligned}$$

since $\mathbb{F}_n(k_m) = \tilde{F}_n(k_m) = 1$ and for any integers $r < s$

$$\begin{aligned}(x-r)_+ - (x-s)_+ &= s - r \quad \text{if } x \geq s \\ &= x - r \quad \text{if } r \leq x < s \\ &= 0 \qquad \text{otherwise.}\end{aligned}$$

Here, j_x is defined to be the unique index such that $k_{j_x} \leq x < k_{j_x+1}$. Hence,

$$\sum_{j=1}^{j_x-1} \mathbb{F}_n(k_j)(k_{j+1} - k_j) + \mathbb{F}_n(k_{j_x})(x - k_{j_x}) \geq \sum_{k=k_1}^{x-1} \tilde{F}_n(k).$$

If x is a knot point, then for $|t|$ small enough the function $\psi + tP_x$ is concave. By Proposition 7.1, it follows that for such an x the condition in (3.4) is satisfied.

Conversely, suppose that $\tilde{\psi}$ satisfies the conditions of Lemma 3.2. Let ψ be a concave function in $\mathcal{C}_m(\mathcal{I})$. By (strict) concavity of Φ_n and using the same summation by parts as above, we have

$$\begin{aligned}\Phi_n(\tilde{\psi}) - \Phi_n(\psi) &\geq \sum_{j=1}^{m} w_j(\tilde{\psi}(k_j) - \psi(k_j)) - \sum_{k=k_1}^{k_m} \exp\tilde{\psi}(k)(\tilde{\psi}(k) - \psi(k)) \\ &= \sum_{j=1}^{m-1} \mathbb{F}_n(k_j)\Big(\tilde{\psi}(k_j) - \psi(k_j) - (\tilde{\psi}(k_{j+1}) - \psi(k_{j+1})\Big) \\ &\quad - \sum_{k=k_1}^{k_m-1} \tilde{F}_n(k)\Big(\tilde{\psi}(k) - \psi(k) - (\tilde{\psi}(k+1) - \psi(k+1)\Big) \\ &= \sum_{j=1}^{m-1} \mathbb{F}_n(k_j)\Big(\tilde{\psi}(k_j) - \tilde{\psi}(k_{j+1})\Big) - \sum_{k=k_1}^{k_m-1} \tilde{F}_n(k)\Big(\tilde{\psi}(k) - \tilde{\psi}(k+1)\Big) \\ &\quad - \left\{\sum_{j=1}^{m-1} \mathbb{F}_n(k_j)\left(\psi(k_j) - \psi(k_{j+1})\right) - \sum_{k=k_1}^{k_m-1} \tilde{F}_n(k)\left(\psi(k) - \psi(k+1)\right)\right\}.\end{aligned}$$

Now, using (3.2), both conditions (3.4) and (3.3) and the fact that $c_i < 0$ we have

$$\sum_{j=1}^{m-1} \mathbb{F}_n(k_j)\left(\psi(k_j) - \psi(k_{j+1})\right) - \sum_{k=k_1}^{k_m-1} \tilde{F}_n(k)\left(\psi(k) - \psi(k+1)\right) \leq 0.$$

Indeed, for $i = 1, ..., p$ we have by the calculations above and taking $x = a_i$

$$\sum_{j=1}^{m-1} \mathbb{F}_n(k_j)\left((a_i - k_j)_+ - (a_i - k_{j+1})_+\right) - \sum_{k=k_1}^{k_m-1} \tilde{F}_n(k)\left((a_i - k)_+ - (a_i - (k+1))_+\right) \geq 0.$$

On the other hand

$$\begin{aligned}&\sum_{j=1}^{m-1} \mathbb{F}_n(k_j)\left((a + bk_j) - (a + bk_{j+1})\right) - \sum_{k=k_1}^{k_m-1} \tilde{F}_n(k)\left(a + bk - (a + b(k+1))\right) \\ &= -b\left\{\sum_{j=1}^{m-1} \mathbb{F}_n(k_j)(k_{j+1} - k_j) - \sum_{k=k_1}^{k_m-1} \tilde{F}_n(k)\right\} = 0\end{aligned}$$

which follows from (3.4) for the knot point k_m. Finally,

$$\sum_{j=1}^{m-1} \mathbb{F}_n(k_j)\left(\tilde{\psi}(k_j)-\tilde{\psi}(k_{j+1})\right)-\sum_{k=k_1}^{k_m-1}\tilde{F}_n(k)\left(\tilde{\psi}(k)-\tilde{\psi}(k+1)\right)=0$$

using (3.2) and the equality condition in (3.4) for the knot points of $\tilde{\psi}$ (including k_m). It follows that $\Phi_n(\tilde{\psi}) \geq \Phi_n(\psi)$, and that $\tilde{\psi}=\widehat{\psi}$. □

Corollary 7.2. *Suppose that κ is a double knot of the log-MLE. Then $\widehat{F}_n(\kappa)=\mathbb{F}_n(\kappa)$. If κ is a triple knot of the log-MLE, then $\widehat{p}_{n,\kappa}=\bar{p}_{n,\kappa}$.*

Proof. From Lemma 3.2, we have that for any knot point κ the equality

$$\sum_{k\leq\kappa-1}\sum_{x\leq k}\widehat{p}_{n,x} \quad = \quad \sum_{k\leq\kappa-1}\sum_{x\leq k}\bar{p}_{n,x}$$

holds. If κ is a double knot, then the above equality holds at both κ and $\kappa+1$. Taking differences, yields that

$$\widehat{F}_n(\kappa)=\sum_{k\leq\kappa}\sum_{x\leq k}\widehat{p}_{n,x}-\sum_{k\leq\kappa-1}\sum_{x\leq k}\widehat{p}_{n,x} \quad = \quad \sum_{k\leq\kappa}\sum_{x\leq k}\bar{p}_{n,x}-\sum_{k\leq\kappa-1}\sum_{x\leq k}\bar{p}_{n,x}=\mathbb{F}_n(\kappa).$$

If κ is a triple knot, then the first equality holds at both $\kappa-1,\kappa$ and $\kappa+1$. Hence,

$$\widehat{p}_{n,\kappa}=\widehat{F}_n(\kappa)-\widehat{F}_n(\kappa-1) \quad = \quad \mathbb{F}_n(\kappa)-\mathbb{F}_n(\kappa-1)=\bar{p}_{n,\kappa}.$$

□

7.6 Proofs from Section 4.1

We begin with the following lemma.

Lemma 7.3. *Let p_n, p denote probability mass functions on $\mathbb{Z}$. The following statements are equivalent.*

- $p_{n,x}\to p_x$ *for all* $x\in\mathbb{Z}$
- $\ell_k(p_n,p)\to 0$ *for any* $1\leq k\leq\infty$
- $\mathcal{H}(p_n,p)\to 0$.

Furthermore, if p_n is log-concave for all n, then p is also log-concave, and the above statements are also equivalent to

- *There exists an $\alpha_0>0$, which depends on the pmf p, such that $\sum_k e^{\alpha|k|}|p_{n,k}-p_k|\to 0$, for all $\alpha\in[0,\alpha_0]$.*

Note that the second part of the lemma is essentially Proposition 2 of Cule and Samworth (2010) in the discrete setting. The proof in our case is, naturally, considerably simpler. We leave the proof of this statement to the Appendix.

Proof of Lemma 7.3. Suppose first that p_n converges to p pointwise. Fix $\varepsilon > 0$. Then there exists a K such that $\sum_{|x|\leq K} p_x \geq 1 - \varepsilon/2$. Furthermore, there exists an n_0 such that $\sup_{|x|\leq K} |p_{n,x} - p_x| \leq \varepsilon/2(2K+1)$, which implies that $\sum_{|x|\leq K} p_{n,x} \geq 1 - \varepsilon$. It follows that

$$\sum_x |p_{n,x} - p_x| \leq \sum_{|x|\leq K} |p_{n,x} - p_x| + \sum_{|x|>K} p_{n,x} + \sum_{|x|>K} p_x < 2\varepsilon.$$

Therefore $p_n \to p$ in ℓ_1. However, since p_n, p are also elements of ℓ_k for all $1 < k \leq \infty$, it also holds that $p_n \to p$ in ℓ_k for any $1 < k \leq \infty$. For convergence in Hellinger distance, ρ_H, we recall that $\rho_H^2(p,q) \leq 2^{-1}\rho_1(p,q)$. Finally, note that convergence in ρ_k or ρ_H implies pointwise convergence. This proves the first part of the lemma.

To prove the second part, recall that if $p_n \to p$ pointwise then there exist random variables such that $X_n \Rightarrow X$ and X_n, X have distributions p_n, p. It is known that log-concavity is preserved under weak limits (see Dharmadhikari and Joag-Dev (1988, Chapter 4)), and therefore p is log-concave. Furthermore, $p_n \to p$ pointwise.
Next, let $\varphi_{n,k} = \log p_{n,k}$ and similarly $\varphi_k = \log p_k$. It follows that $\varphi_{n,k} \to \varphi_k$ pointwise, and therefore also uniformly for k bounded. Without loss of generality, assume that 0 is in the support of the mass function p. Also, there exist $b \in \mathbb{R}$ and $c > 0$ such that $p_k \leq e^{-c|k|+b}$ for all $k \in \mathbb{Z}$. Let $\alpha_0 = c$ and define $\varepsilon = (\alpha_0 - \alpha)/4 > 0$. Then

$$\frac{\varphi_k - \varphi_0}{|k|} \leq -\alpha_0 + (b - \varphi_0)/|k| \leq -\alpha_0 + \varepsilon$$

for all $|k| \geq R$ for some sufficiently large R.
Since φ_n is log-concave, it follows that $(\varphi_{n,k} - \varphi_{n,0})/|k|$ is decreasing in k for $|k| \geq R$, as long as R was chosen large enough. Hence, we also have that

$$\frac{\varphi_{n,k} - \varphi_{n,0}}{|k|} \leq -\alpha_0 + 2\varepsilon$$

for all $|k| \geq R$ and all n sufficiently large. We have thus shown that

$$\begin{aligned}\sum_k e^{\alpha|k|} p_{n,k} &\leq \sum_{|k|<R} e^{\alpha|k|} p_{n,k} + \sum_{|k|\geq R} e^{\alpha|k| + \varphi_{n,0} + (-\alpha_0 + 2\varepsilon)|k|} \\ &\leq \sum_{|k|<R} e^{\alpha|k|} p_{n,k} + e^{\varphi_{n,0}} \sum_{|k|\geq R} e^{-2\varepsilon|k|},\end{aligned}$$

from which it follows that $\sum_k e^{\alpha|k|} p_{n,k}$ is summable. The result now follows from the dominated convergence theorem. $\square$

Proof of Theorem 4.1. The proof is divided into several steps. First, consider the pmf $\widetilde{q}$ proportional to $e^{-|k|}$. Then $\rho_{\mathrm{KL}}(\widetilde{q}\,\|\,p)$ is finite by the assumptions of the theorem, and hence there exists a sequence of log-concave probability mass functions q_n such that $\rho_{\mathrm{KL}}(q_n\,\|\,p) \to \operatorname{argmin}_{q\in\mathcal{LC}_1} \rho_{\mathrm{KL}}(q\,\|\,p)$. Since $\rho_{\mathrm{KL}}(q_n\,\|\,p) \leq \rho_{\mathrm{KL}}(\widetilde{q}\,\|\,p)$, it follows that

$$\sup_n \sum_{x\in\mathbb{Z}} |-\log q_{n,x}| p_x \quad \leq \quad \sum_{x\in\mathbb{Z}} (-\log \widetilde{q}_x) p_x < \infty,$$

which in turn implies that for any fixed m

$$\sup_n \sum_{|x|\leq m} |-\log q_{n,x}| \quad \leq \quad \sup_{|x|\leq m} \frac{1}{p_x} \sum_{x\in\mathbb{Z}} (-\log \widetilde{q}_x) p_x < \infty.$$

Therefore, for each m, there exists a $\delta > 0$ such that $\inf_n \inf_{|x|\leq m} q_{n,x} \geq \delta$. If M is sufficiently large, it follows that $\limsup_n \sup_{|x|>M} q_{n,x} \leq \delta/2$, which in turn implies that there exists a function $h_x = -\alpha|x| + \beta$ such that $\sup_n q_{n,x} \leq e^{h_x}$.

Let X_n denote the random variable with pmf q_n. It follows from the above that X_n is tight, and therefore q_n has a convergent subsequence. Let q_0 denote its log-concave limit. By Fatou's lemma

$$\rho_{\mathrm{KL}}\left(q_0 \,\|\, p\right) \quad \leq \quad \liminf_n \rho_{\mathrm{KL}}\left(q_n \,\|\, p\right) = \mathrm{argmin}_{q\in\mathcal{LC}_1} \rho_{\mathrm{KL}}\left(q \,\|\, p\right),$$

which shows that a minimiser exists.

Next, suppose that $\widehat{p}_1$ and $\widehat{p}_2$ both minimise $\rho_{\mathrm{KL}}\left(\cdot \,\|\, p\right)$, and let $\widetilde{p}$ be the pmf proportional to $(\widehat{p}_1\widehat{p}_2)^{1/2}$. Then

$$\sum p \log \frac{p}{\widetilde{p}} = \sum p \log \frac{p}{\widehat{p}_1} + \log \sum (\widehat{p}_1\widehat{p}_2)^{1/2} \leq \sum p \log \frac{p}{\widehat{p}_1}$$

by the Cauchy-Schwarz inequality, with equality if and only if $\widehat{p}_1 = \widehat{p}_2$. Therefore, the minimiser is also unique.

It remains to show that $\widehat{p}_n \to \widehat{p}$. Recall that $\widehat{p}_n$ satisfies the inequality in (3.6). Let X_n denote a random variable with pmf $\widehat{p}$. Then, by Markov's inequality,

$$P(|X_n| \geq m) \leq \frac{\sum_{x\in\mathbb{Z}} |x| \bar{p}_{n,x}}{m}.$$

Therefore, X_n is tight, and hence there exists a subsequence of $\widehat{p}_n$, which we denote again by $\{n\}$, such that $\widehat{p}_n \to \widetilde{p}$, for some pmf $\widetilde{p}$. From Lemma 7.3 it follows that $\widetilde{p}$ must also be log-concave. It remains to show that $\widetilde{p} = \widehat{p}$ to complete the proof.
By definition of the MLE, we have that for any $b > 0$

$$\begin{aligned} 0 \quad &\leq \quad \sum \log(b + \widehat{p}_{n,k})\, \bar{p}_{n,k} - \sum \log \widehat{p}_k \, \bar{p}_{n,k} \\ &= \quad \sum \log(b + \widehat{p}_{n,k})\, (\bar{p}_{n,k} - p_k) + \sum \log \widehat{p}_k \, (p_k - \bar{p}_{n,k}) \\ &\qquad\qquad + \sum \log \left(\frac{b + \widehat{p}_{n,k}}{b + \widehat{p}_k} \right) p_k + \sum \log \left(\frac{b + \widehat{p}_k}{\widehat{p}_k} \right) p_k \end{aligned}$$

We next get rid of the first two term on the right-hand side. First, using summation by parts

$$\sum_{k\in\mathbb{Z}} \log\left(b + \widehat{p}_{n,k}\right)(\bar{p}_{n,k} - p_k) \quad = \quad \sum_{k\in\mathbb{Z}} (\mathbb{F}_n(k) - F(k)) \left[\log\left(b + \widehat{p}_{n,k}\right) - \log\left(b + \widehat{p}_{n,k-1}\right)\right],$$

and hence

$$\begin{aligned}
&\left|\sum_{k\in\mathbb{Z}} \log\left(b+\widehat{p}_{n,k}\right)\left(\bar{p}_{n,k}-p_k\right)\right| \\
&\qquad \le \sup_{k\in\mathbb{Z}} |\mathbb{F}_n(k)-F(k)| \left\{ \sum_{k\le \widehat{m}} \left[\log\left(b+\widehat{p}_{n,k}\right)-\log\left(b+\widehat{p}_{n,k-1}\right)\right] \right. \\
&\qquad\qquad\qquad \left. + \sum_{k>\widehat{m}} \left[\log\left(b+\widehat{p}_{n,k}\right)-\log\left(b+\widehat{p}_{n,k-1}\right)\right] \right\} \\
&\qquad \le 2\log(b+\widehat{p}_{n,\widehat{m}}) \sup_{k\in\mathbb{Z}} |\mathbb{F}_n(k)-F(k)| \\
&\qquad \le 2\log(1+b) \sup_{k\in\mathbb{Z}} |\mathbb{F}_n(k)-F(k)|,
\end{aligned}$$

which converges to zero. The law of large numbers shows that the second term also converges to zero. Therefore,

$$\limsup_n \sum \log\left(\frac{b+\widehat{p}_k}{b+\widehat{p}_{n,k}}\right) p_k \quad \le \quad \sum \log\left(\frac{\widehat{p}_k}{b+\widehat{p}_k}\right) p_k$$

which implies that (using Fatou's lemma)

$$\limsup_{b\to 0} \limsup_n \sum \log\left(\frac{b+\widehat{p}_k}{b+\widehat{p}_{n,k}}\right) p_k \quad \le \quad 0$$

Taking limits (by the dominated and monotone convergence theorems) yields that

$$\sum \log\left(\frac{\widehat{p}_k}{\widetilde{p}_k}\right) p_k \quad \le \quad 0,$$

or, in other words,

$$\sum \log\left(\frac{p_k}{\widetilde{p}_k}\right) p_k \quad \le \quad \sum \log\left(\frac{p_k}{\widehat{p}_k}\right) p_k,$$

which, as $\widehat{p}$ is the unique minimiser of the quantity on the right hand side, proves that $\widehat{p} = \widetilde{p}$. □

Proof of Corollary 4.2. This follows since

$$|\widehat{F}_n(x) - \widehat{F}(x)| \quad \le \quad \sum_{y\le x} |\widehat{p}_{n,y} - \widehat{p}_y| \quad \le \quad \ell_1(\widehat{p}_n, \widehat{p}).$$

□

Proof of Lemma 4.3. A point r is a knot of $\widehat{\psi}_n$ if, and only if, $(\Delta\widehat{\psi}_n)_s < 0$. Let $\varepsilon = -(\Delta\psi)_x > 0$. It follows from Proposition 4.1, that there exists an n_0 such that for all $n \ge n_0$

$$\max_{x=r-1,r,r+1} |\widehat{\psi}_{n,x} - \psi_x| < \varepsilon/8,$$

with probability one. Therefore,

$$\begin{aligned}(\Delta\widehat{\psi}_n)_r &= (\Delta\psi)_r + (\Delta[\widehat{\psi}_n - \psi])_r \\ &\leq (\Delta\psi)_r + 4\left(\max_{x=r-1,r,r+1} |\widehat{\psi}_{n,x} - \psi_x|\right) \\ &< -\varepsilon + \varepsilon/2,\end{aligned}$$

which is strictly negative, as required. □

7.7 Proofs from Section 4.2

The following proposition is crucial for establishing the weak convergence of the MLE.

Proposition 7.4. *Let $r < s$ denote two successive knot points. Then for all $x \in \{r, \ldots, s-1\}$, $\sqrt{n}\ (\widehat{\psi}_{n,x} - \psi_x)$ and $\sqrt{n}\ (\widehat{p}_{n,x} - p_x)$ are bounded in probability for all n.*

To prove the above proposition, we start with the following lemma.

Lemma 7.5. *If r is a knot of $\widehat{\psi}_n$ then*

$$\bar{p}_{n,r} \geq \widehat{p}_{n,r}, \tag{7.6}$$

Proof. Let q denote the first knot point before r. From Lemma 3.2 we have the following statements

$$\sum_{x=q}^{r} \mathbb{F}_n(x) \geq \sum_{x=q}^{r} \widehat{F}_n(x) \tag{7.7}$$

$$\sum_{x=q}^{r-1} \mathbb{F}_n(x) = \sum_{x=q}^{r-1} \widehat{F}_n(x) \tag{7.8}$$

$$\sum_{x=q}^{r-2} \mathbb{F}_n(x) \geq \sum_{x=q}^{r-2} \widehat{F}_n(x), \tag{7.9}$$

where $\widehat{F}_n$ denotes the cdf of the MLE. Taking (7.7) − 2(7.8) + (7.9) yields (7.6). □

Proof of Proposition 7.4. Showing boundedness in probability of $\sqrt{n}(\widehat{p}_n - p)$ on $\{r, \ldots, s-1\}$ for any consecutive knots of $\psi = \log p$ is equivalent to showing the same property on $\{0, \ldots, r-1\}$ for any knot r occuring the left of zero. If $r = 1$, then Lemma 3.2 implies that $\bar{p}_{n,0} = \widehat{p}_{n,0}$ and boundedness immediately follows at $0 = r - 1$. From now on, $r > 1$ will be assumed. Let $0 \leq u < v \leq r$ be consecutive (random) knots of $\widehat{\psi}_n$ such that $\{u, \ldots, v\}$ is the first (random) segment containing at least one interior point; i.e., $v > u + 1$. Then, all knots occuring before u are triple knots and by Corollary 7.2 it follows that

$$\sup_{0\leq k\leq u-1} \sqrt{n}|\widehat{p}_{n,k} - p_k| = \sup_{0\leq k\leq u-1} \sqrt{n}|\bar{p}_{n,k} - p_k| \leq \sup_{0\leq k\leq r-1} \sqrt{n}|\bar{p}_{n,k} - p_k| = O_p(1).$$

We will now show that $\sup_{u\le k\le v-1}\sqrt{n}|\widehat{p}_{n,k}-p_k| = O_p(1)$. Let $\delta > 0$. For a given $x \in \{u,\dots,v-1\}$, consider the perturbation function $\alpha_{u,v,x}$, where

$$\alpha_{u,v,x}(k) = \begin{cases} \frac{k-u}{x-u}, & \text{if } k \in \{u,\dots,x\} \\ \frac{v-u}{v-x}, & \text{if } k \in \{x+1,\dots,v-1\}. \end{cases}$$

If δ is chosen to be small enough, then it is not difficult to see that $\widehat{\psi} + \delta\alpha_{u,v,x} \in \mathcal{C}_m(\mathcal{I})$. Applying Proposition 7.1 yields

$$\sum_{k=u+1}^{v-1} \bar{p}_{n,k}\,\alpha_{u,v,x}(k) \le \sum_{k=u+1}^{v-1} \widehat{p}_{n,k}\,\alpha_{u,v,x}(k).$$

Using Lemma 7.5 and the fact that both ψ and $\widehat{\psi}_n$ are linear on $\{u,\dots,v\}$, we can write for $k \in \{u+1,\dots,v-1\}$

$$p_k = p_u^{\frac{v-k}{v-u}} p_v^{\frac{k-u}{v-u}} \qquad \text{and} \qquad \widehat{p}_{n,k} = \widehat{p}_{n,u}^{\frac{v-k}{v-u}} \widehat{p}_{n,v}^{\frac{k-u}{v-u}} \le \bar{p}_{n,u}^{\frac{v-k}{v-u}} \bar{p}_{n,v}^{\frac{k-u}{v-u}}. \tag{7.10}$$

It follows that

$$\begin{aligned} \sum_{k=u+1}^{v-1} \sqrt{n}(\bar{p}_{n,k}-p_k)\alpha_{u,v,x,k} &\le \sum_{k=u+1}^{v-1} \sqrt{n}(\widehat{p}_{n,k}-p_k)\alpha_{u,v,x,k} \\ &\le \sum_{k=u+1}^{v-1} \sqrt{n}\left(\bar{p}_{n,u}^{\frac{v-k}{v-u}} \bar{p}_{n,v}^{\frac{k-u}{v-u}} - p_u^{\frac{v-k}{v-u}} p_v^{\frac{k-u}{v-u}}\right)\alpha_{u,v,x}(k). \end{aligned}$$

Since $\alpha_{u,v,x} \le 1$ this further implies

$$\begin{aligned} -(r-1)\inf_{0\le k\le r}\sqrt{n}|\bar{p}_{n,k}-p_k| &\le \sum_{k=u+1}^{v-1} \sqrt{n}(\widehat{p}_{n,k}-p_k)\alpha_{u,v,x,k} \\ &\le (r-1)\sup_{0\le u<v\le r, u+1\le k\le v-1}\sqrt{n}\left|\bar{p}_{n,u}^{\frac{v-k}{v-u}} \bar{p}_{n,v}^{\frac{k-u}{v-u}} - p_u^{\frac{v-k}{v-u}} p_v^{\frac{k-u}{v-u}}\right|. \end{aligned}$$

Using the central limit theorem and the delta method, it follows that

$$\sum_{k=u+1}^{v-1} \sqrt{n}(\widehat{p}_{n,k}-p_k)\alpha_{u,v,x}(k) = O_p(1) \tag{7.11}$$

for all $x \in \{u+1,\dots,v-1\}$, and where $O_p(1)$ does not depend on u, v nor x. Hence, we have obtained a linear system, and (7.11) can be rewritten as

$$A\left(\sqrt{n}(\widehat{p}_{n,u+1}-p_{u+1})\dots\sqrt{n}(\widehat{p}_{n,v-1}-p_{v-1})\right)^t = (O_p(1)\dots O_p(1))^t$$

where A is the $(v-u-1)\times(v-u-1)$ matrix $(\alpha_{u,v,x}(k))_{u+1\le k,x\le v-1}$. By some algebra we can show that A is invertible such that

$$A^{-1} = \begin{pmatrix} \frac{2(v-u-1)}{v-u} & -\frac{2(v-u-2)}{v-u} & 0 & 0 & 0 & 0 & \dots & 0 \\ -\frac{v-u-1}{v-u} & \frac{4(v-u-2)}{v-u} & -\frac{3(v-u-3)}{v-u} & 0 & 0 & 0 & \dots & 0 \\ 0 & -\frac{2(v-u-2)}{v-u} & \frac{6(v-u-3)}{v-u} & -\frac{4(v-u-4)}{v-u} & 0 & 0 & \dots & 0 \\ 0 & 0 & -\frac{3(v-u-3)}{v-u} & \frac{8(v-u-4)}{v-u} & -\frac{5(v-u-5)}{v-u} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \dots & 0 & -\frac{2(v-u-2)}{v-u} & \frac{2(v-u-1)}{v-u} \end{pmatrix}.$$

Since $(v-u-1)/(v-u) \leq r-1$, A^{-1} admits bounded entries which in turn implies that the $\sqrt{n}(\widehat{p}_{n,k} - p_k) = O_p(1)$ for all $k \in \{u+1, \dots, v-1\}$. If $v > u+2$, then boundedness in probability at u (and also at v) follows from the previous conclusion and applying the delta method on any pair of the identities given in (7.10). If $v = u+2$, then by Lemma 3.2 we have that

$$\sum_{k=u}^{v-1} \mathbb{F}_n(k) = \sum_{k=u}^{v-1} \widehat{F}_n(k)$$

or equivalently

$$\sqrt{n}(\widehat{p}_{n,u} - p_u) \quad = \quad \sqrt{n}(\bar{p}_{n,u} - p_u) + \frac{1}{2}\left[\sqrt{n}(\bar{p}_{n,u+1} - p_{u+1}) - \sqrt{n}(\widehat{p}_{n,u+1} - p_{u+1})\right]$$

and boundedness at u follows by the central limit theorem and boundedness at $u+1$ as shown above. The argument above can be repeated on the subsequent segments. Finally, boundedness of $\sqrt{n}(\widehat{\psi}_n - \psi)$ follows by applying the delta method. $\square$

Proof of Proposition 4.4. Existence of a minimiser follows from arguments along the same lines as in the proof of Theorem 3.1-(ii) since the functions $x \mapsto -\left(x^2/2 - x\ \mathbb{W}(x)\right)$ are coercive for $k = r, \dots, s-1$. Uniqueness is a result of strict convexity of Φ.

Now, fix $x \in \{r+1, \dots, s-1\}$. For $t > 0$, the function $k \mapsto g^*(k) + tP_x(k)$, where P_x is the perturbation function defined in (7.5), is concave and hence

$$0 \leq \lim_{t \searrow 0} \frac{\Phi(g^* + tP_x) - \Phi(g^*)}{t} \quad = \quad -\sum_{k=r}^{s-1} (g^*(k)p_k - \mathbb{W}(k))P_x(k).$$

Repeating the same calculation as in the proof of Lemma 3.2 shows that this is equivalent to

$$\sum_{k=r}^{x-1} \sum_{y=r}^{k} (g^*(y)p_y - \mathbb{W}(y)) \leq 0,$$

which shows the inequality part of (4.2). Proof of equality if x is a knot of g^* uses the fact that the perturbation function P_x satisfies that $g^* \pm tP_x$ is concave for $|t|$ small enough yielding $\lim_{t \to 0}(\Phi(g^* - tP_x) - \Phi(g^*))/t = 0$. In this sense, s is always a knot point since the function P_s is simply the linear function on $\{r, r+1, \dots, s-1\}$. Therefore we always have equality at s, and also trivially at $x = r$ by definition. It follows that the minimiser of Φ must satisfy the conditions (4.2).

Conversely, suppose that g^* is a concave function on $\{r, \dots, s-1\}$ such that the process $\mathbb{H}$ defined in (4.3) satisfies (4.2). Let g be a concave function on $\{r, \dots, s-1\}$. We will show

that $\Phi(g) \geq \Phi(g^*)$. We have

$$\begin{aligned}
\Phi(g) - \Phi(g^*) &= \frac{1}{2}\sum_{k=r}^{s-1}\left[g^2(k) - (g^*)^2(k)\right] p_k - \sum_{k=r}^{s-1}\left[g(k) - g^*(k)\right] \mathbb{W}(k) \\
&= \frac{1}{2}\sum_{k=r}^{s-1}(g(k) - g^*(k))^2 p_k + \sum_{k=r}^{s-1}(g(k) - g^*(k))\left\{g^*(k)\, p_k - \mathbb{W}(k)\right\} \\
&\geq \sum_{k=r}^{s-1}(g(k) - g^*(k))\left\{g^*(k)\, p_k - \mathbb{W}(k)\right\}.
\end{aligned}$$

Now, for $a \in \{r, \dots, s\}$

$$\sum_{k=r}^{s-1}\left\{g^*(k)\, p_k - \mathbb{W}(k)\right\}(a-k)_+ = \sum_{k=r}^{a-1}\sum_{y=r}^{k}\left\{g^*(k)\, p_k - \mathbb{W}(k)\right\} = \mathbb{H}(a) - \mathbb{Y}(a) \leq 0,$$

with equality if a is knot of g^*, by assumption. Using a representation for concave functions on $\in \{r, \dots, s-1\}$ similar to (3.2) (note that the functions $(a-k)_+$ have negative weights), it follows that

$$\sum_{k=r}^{s-1} g(k)\left\{g^*(k)\, p_k - \mathbb{W}(k)\right\} \quad \geq \quad 0,$$

with equality if g is replaced with g^*. It follows that $\Phi(g) \geq \Phi(g^*)$ and hence that g^* is the minimiser of Φ. □

Lemma 7.6. *The limiting distribution $\mathbb{H}$ also satisfies $\sum_{x=r}^{s-1} g^*(x)p_x = \sum_{x=r}^{s-1}\mathbb{W}(x)$.*

Proof. The proof follows from the argument above, choosing the perturbation function $P(k) = \pm 1$. □

Proof of Theorem 4.5. From the definition of the process and Proposition 7.4, it follows that $\widehat{\mathbb{H}}_n$ is tight. Therefore, we apply Prokhorov's theorem, and it is sufficient to show that any convergent subsequence has the same limit, $\mathbb{H}$. To do this, consider a subsequence of $\widehat{\mathbb{H}}_n$ (which we denote again by $\{n\}$,) and let $\mathbb{Q}$ denote its weak limit. Note that all we need to do is to show that $\mathbb{Q}$ satisfies the condition in (4.2) with equality at $x = r, s$, as these uniquely identify this distribution.

We first prove the result on the first segment $\{0, \dots, s\}$. Lemma 3.2 implies that

$$\sum_{k=0}^{y-1}\sum_{x=0}^{k}\widehat{p}_{n,x} \quad \leq \quad \sum_{k=0}^{y-1}\sum_{x=0}^{k}\bar{p}_{n,x},$$

with equality at all knots y, where s is always a knot. Subtracting from both sides $\sum_{k=0}^{y-1}\sum_{x=0}^{k} p_x$ and multiplying by $\sqrt{n}$, yields that $\widehat{\mathbb{H}}_n(y) \leq \mathbb{Y}_n(y)$ with equality at all knots y. Note that in

the definition of $\mathbb{Y}_n$ and $\widehat{\mathbb{H}}_n$ given above (and that of the related processes $\mathbb{G}_n$, $\mathbb{W}_n$, $\widehat{\mathbb{G}}_n$ and $\widehat{\mathbb{W}}_n$) the knot r should be replaced by 0. Hence,

$$\begin{aligned}\mathbb{Y}_n(y) \geq \widehat{\mathbb{H}}_n(y) &= \sum_{k=0}^{y-1}\sum_{x=0}^{k}\sqrt{n}(\widehat{p}_{n,x}-p_x) = \sum_{k=0}^{y-1}\sum_{x=0}^{k}\sqrt{n}(\exp\widehat{\psi}_{n,x}-\exp\psi_x)\\ &= \sum_{k=0}^{y-1}\sum_{x=0}^{k}p_x\sqrt{n}(\widehat{\psi}_{n,x}-\psi_x)+\sum_{k=0}^{y-1}\sum_{x=0}^{k}\exp\theta_{n,x}\sqrt{n}(\widehat{\psi}_{n,x}-\psi_x)^2/2,\end{aligned}$$

for some $\theta_{n,y}$, by applying a Taylor series expansion of second order. The inequality holds for all $x \in \{0,\ldots,s\}$ with equality at knots of $\widehat{\psi}_n$. Note also that since ψ is linear on $\{0,\ldots,s\}$, the knots of $\widehat{\psi}_n$ are the same as the knots of $\sqrt{n}(\widehat{\psi}_n-\psi)$. Let $\widetilde{\mathbb{H}}_n(y) = \sum_{k=0}^{y-1}\sum_{x=0}^{k}p_x\sqrt{n}(\widehat{\psi}_{n,x}-\psi_x)$. Then, by Proposition 7.4,

$$\mathbb{Y}_n(x)+o_p(1) \quad\geq\quad \widetilde{\mathbb{H}}_n(x)$$

for all $x=0,\ldots,s$ with equality at all knots of $\sqrt{n}(\widehat{\psi}_n-\psi)$ on $\{0,\ldots,s-1\}$. Note also that $\sqrt{n}(\widehat{\psi}_n-\psi)$ is concave on $\{0,\ldots,s-1\}$.

It is well-known that $\mathbb{Y}_n(y)$ converges in distribution to $\mathbb{Y}(y)$ for all $y=0,\ldots,s$, with $\mathbb{Y}\in\mathbb{R}^{\{r,\ldots,s\}}$ defined in Proposition 4.4 (for any successive knots $r<s$), and therefore so does the process $\mathbb{Y}_n(x)+o_p(1)$. Recall that $\mathbb{Q}$ denotes the weak limit of $\widehat{\mathbb{H}}_n$, which is the same as the weak limit of $\widetilde{\mathbb{H}}_n=\widehat{\mathbb{H}}_n+o_p(1)$. Taking limits in the last display above thus shows that

$$\mathbb{Y}(x)\geq\mathbb{Q}(x),$$

with equality at $x=0,s$ and at all knot points of the process $(\Delta\mathbb{Q})_x/p_x$. The latter follows from the fact that such a knot point is also a knot point of

$$\frac{(\Delta\widetilde{\mathbb{H}}_n)_x}{p_x} \quad=\quad \sqrt{n}(\widehat{\psi}_{n,x}-\psi).$$

Therefore, $\mathbb{Q}$ satisfies the conditions in (4.2), and the same boundary conditions at $x=0,s$ as $\mathbb{H}$. Hence, $\mathbb{Q}\stackrel{d}{=}\mathbb{H}$, as required.

Lastly, note that

$$\begin{aligned}\sqrt{n}(\widehat{p}_{n,x}-p_x) &= (\Delta\mathbb{H}_n)_x,\\ \text{and}\quad \sqrt{n}(\widehat{F}_n(x)-F(x)) &= (\nabla\mathbb{H}_n)_x\end{aligned}$$

since $\widehat{F}_n(-1)=\mathbb{F}_n(-1)=0$. This completes the proof on the segment $\{0,\ldots,s\}$.

Now, let $s'>s$ be the next knot of the true log-pmf. The argument here is the same, except that we we no longer have that $\widehat{F}_n(s-1)=\mathbb{F}_n(s-1)$ as we did for the knot $x=0$. However, it is true that $\sqrt{n}(\widehat{F}_n(s-1)-\mathbb{F}_n(s-1))\to_p 0$. This follows since

$$\begin{aligned}\sqrt{n}(\widehat{F}_n(s-1)-\mathbb{F}_n(s-1)) &= \sqrt{n}(\widehat{F}_n(s-1)-F(s-1))-\sqrt{n}(\mathbb{F}_n(s-1)-F(s-1))\\ &\to_d \mathbb{H}(s)-\mathbb{H}(s-1)-(\mathbb{Y}(s)-\mathbb{Y}(s-1))\\ &= \sum_{x=r}^{s-1}g^*(x)p_x-\sum_{x=r}^{s-1}\mathbb{W}(x)=0,\end{aligned}$$

by Lemma 7.6. This allows us to repeat the above argument on the segment $\{s,\ldots,s'-1\}$, and so on. Iterating in this fashion establishes the result in the general case. □

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Alzaid, A. A. and Omair, M. A. (2010). On the Poisson difference distribution inference and applications. *Bull. Malays. Math. Sci. Soc. (2)* **33** 17–45.

Asgharian, M. and Wolfson, D. (2002). Length-biased sampling with right censoring: An unconditional approach. *Journal of the American Statistical Association* **97** 201–209.

Balabdaoui, F., Rufibach, K. and Wellner, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331.

Banerjee, M. and Wellner, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731.

Bardwell, G. E. and Crow, E. L. (1964). A two-parameter family of hyper-Poisson distributions. *J. Amer. Statist. Assoc.* **59** 133–141.

Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* **25** 970–981.

Carolan, C. and Dykstra, R. (1999). Asymptotic behavior of the Grenander estimator at density flat regions. *Canad. J. Statist.* **27** 557–566.

Crow, E. L. and Bardwell, G. E. (1965). Estimation of the parameters of the hyper-Poisson distributions. In *Classical and Contagious Discrete Distributions (Proc. Internat. Sympos., McGill Univ., Montreal, Que., 1963)*. Statistical Publishing Society, Calcutta, 127–140.

Cule, M. and Samworth, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electronic J. Stat.* **4** 254–270.

Cule, M., Samworth, R. and Stewart, M. (2010). Maximum likelihood estimation of a multidimensional log-concave density. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 545–607.

Devroye, L. (1987). A simple generator for discrete log-concave distributions. *Computing* **39** 87–91.

Dharmadhikari, S. and Joag-Dev, K. (1988). *Unimodality, convexity, and applications.* Probability and Mathematical Statistics, Academic Press Inc., Boston, MA.

Dümbgen, L., Hüsler, A. and Rufibach, K. (2010). Active set and EM algorithms for log-concave densities based on complete and censored data. Tech. rep., University of Bern. Available at arXiv:0707.4643.

Dümbgen, L. and Rufibach, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function. *Bernoulli* **15** 40–68.

DÜMBGEN, L. and RUFIBACH, K. (2011). logcondens: Computations related to univariate log-concave density estimation. *Journal of Statistical Software* **39** 1–28.

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer, Vol. II (Berkeley, Calif., 1983)*. Wadsworth Statist./Probab. Ser., Wadsworth, Belmont, CA.

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001a). A canonical process for estimation of convex functions: the "invelope" of integrated Brownian motion $+t^4$. *Ann. Statist.* **29** 1620–1652.

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001b). Estimation of a convex function: characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698.

JANKOWSKI, H. K. and WELLNER, J. A. (2009). Estimation of a discrete monotone distribution. *Electron. J. Stat.* **3** 1567–1605.

JOHNSON, N. L. and KOTZ, S. (1969). *Distributions in statistics: Discrete distributions*. Houghton Mifflin Co., Boston, Mass.

KARLIS, D. and NTZOUFRAS, I. (2005). Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *Journal of Statistical Software* **14**.

KARLIS, D. and NTZOUFRAS, I. (2006). Bayesian analysis of the differences of count data. *Stat. Med.* **25** 1885–1905.

KARLIS, D. and NTZOUFRAS, I. (2009). Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. *IMA Journal of Management Mathematics* **20** 133–145.

KARLIS, D. and NTZOUFRAS, L. (2003). Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society, Series D* **52** 381–393.

KEILSON, J. and GERBER, H. (1971). Some results for discrete unimodality. *Journal of the American Statistical Association* **66** 386–389.

LEWIS, J. W. (2009). *skellam: Skellam distribution*. R package version 0.0-8-7.
URL `http://CRAN.R-project.org/package=skellam`

NG, P. T. and MAECHLER, M. (2011). *cobs: COBS – Constrained B-splines (Sparse matrix based)*. R package version 1.2-2.
URL `http://CRAN.R-project.org/package=cobs`

R DEVELOPMENT CORE TEAM (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL `http://www.R-project.org/`

ROCKAFELLAR, R. T. (1970). *Convex analysis*. Princeton Mathematical Series, No. 28, Princeton University Press.

RUFIBACH, K. (2006). *Log-concave density estimation and bump hunting for I.I.D. observations*. Ph.D. thesis, Universities of Bern and Göttingen.

RUFIBACH, K., BALABDAOUI, F., JANKOWSKI, H. and WEYERMANN, K. (2011). *logcondiscr: Estimate a Log-Concave Probability Density from Discrete i.i.d. Observations*. R package version 1.0.0.

SEREGIN, A. and WELLNER, J. A. (2010). Nonparametric Estimation of Multivariate Convex-Transformed Densities. *Ann. Statist.* **38** 3751–3781.

SILVERMAN, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.* **10** 795–810.

SKELLAM, J. G. (1946). The frequency distribution of the difference between two Poisson variates belonging to different populations. *J. Roy. Statist. Soc. (N.S.)* **109** 296.

STANLEY, R. P. (1989). Log-concave and unimodal sequences in algebra, combinatorics, and geometry. In *Graph theory and its applications: East and West (Jinan, 1986)*, vol. 576 of *Ann. New York Acad. Sci.* New York Acad. Sci., 500–535.

TANG, R., BANERJEE, M. and KOSOROK, M. (2011). Asymptotics for current status data under varying observation time sparsity. *Preprint* to appear.

TUITE, A. R., GREER, A. L., WHELAN, M., WINTER, A.-L., LEE, B., YAN, P., WU, J., MOGHADAS, S., BUCKERIDGE, D., POURBOHLOUL, B. and FISMAN, D. N. (2010). Estimated epidemiologic parameters and morbidity associated with pandemic H1N1 influenza. *CMAJ* **182** 131–136.

VARADHAN, S. R. S. (1968). *Stochastic Processes*. Notes based on a course given at New York University during the year 1967/68, Courant Institute of Mathematical Sciences New York University, New York.

WALTHER, G. (2009). Inference and modeling with log-concave distributions. *Statist. Sci.* **24** 319–327.

WEYERMANN, K. (2008). *An active-set algorithm for the estimation of discrete log-concave densities*. Master's thesis, University of Bern. In German.